The Journal of Economic Perspectives

A journal of the
American Economic Association

Summer 2020

The Journal of
Economic Perspectives

A journal of the American Economic Association

Editor
Enrico Moretti, University of California, Berkeley

Coeditors
Gordon Hanson, Harvard University
Heidi Williams, Stanford University

Associate Editors
Leah Boustan, Princeton University
Gabriel Chodorow-Reich, Harvard University
Dora Costa, University of California, Los Angeles
Janice Eberly, Northwestern University
David Figlio, Northwestern University
Eliana La Ferrara, Bocconi University
Camille Landais, London School of Economics
Amanda Pallais, Harvard University
Fiona Scott Morton, Yale University
Charlie Sprenger, University of California, San Diego
Gianluca Violante, Princeton University
Ebonya Washington, Yale University
Luigi Zingales, University of Chicago

Managing Editor
Timothy Taylor

Assistant Managing Editor
Alexandra Szczupak

Editorial offices:
Journal of Economic Perspectives
American Economic Association Publications
2403 Sidney St., #260
Pittsburgh, PA 15203
email: jep@aeapubs.org

The *Journal of Economic Perspectives* gratefully acknowledges the support of Macalester College.
Registered in the US Patent and Trademark Office (®).
Copyright © 2020 by the American Economic Association; All Rights Reserved.
Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.
Printed by LSC Communications, Owensville, Missouri, 65066, USA.
No responsibility for the views expressed by the authors in this journal is assumed by the editors or by the American Economic Association.
THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Summer 2020, Vol. 34, No. 3. The JEP is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. Annual dues for regular membership are $24.00, $34.00, or $44.00 depending on income; for an additional $15.00, you can receive this journal in print. The journal is freely available online. For details and further information on the AEA go to https://www.aeaweb.org/. Periodicals postage paid at Nashville, TN, and at additional mailing offices.
POSTMASTER: Send address changes to the *Journal of Economic Perspectives*, 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the U.S.A.

The Journal of
Economic Perspectives

Contents

Volume 34 • Number 3 • Summer 2020

Statement of Purpose

The *Journal of Economic Perspectives* attempts to fill a gap between the general interest press and most other academic economics journals. The journal aims to publish articles that will serve several goals: to synthesize and integrate lessons learned from active lines of economic research; to provide economic analysis of public policy issues; to encourage cross-fertilization of ideas among the fields of economics; to offer readers an accessible source for state-of-the-art economic thinking; to suggest directions for future research; to provide insights and readings for classroom use; and to address issues relating to the economics profession. Articles appearing in the journal are normally solicited by the editors and associate editors. Proposals for topics and authors should be directed to the journal office, at the address inside the front cover.

Policy on Data Availability

It is the policy of the *Journal of Economic Perspectives* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the above requirements cannot be met.

Policy on Disclosure

Authors of articles appearing in the *Journal of Economic Perspectives* are expected to disclose any potential conflicts of interest that may arise from their consulting activities, financial interests, or other nonacademic activities.

Journal of Economic Perspectives

Advisory Board

Stephanie Aaronson, Brookings Institution
Janet Currie, Princeton University
Karen Dynan, Harvard University
Claudia Goldin, Harvard University
Peter Henry, New York University
Kenneth Kuttner, Williams College
Trevon Logan, Ohio State University
David Sappington, University of Florida
Dan Sichel, Wellesley College
Jonathan Skinner, Dartmouth College
Ludger Woessmann, Ifo Institute for Economic Research

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 3–26

The Economics of Urban Density

Gilles Duranton and Diego Puga

Everybody loves density. Economists like to model and quantify the many benefits of urban density. It boosts productivity and innovation, improves access to goods and services, reduces travel needs, encourages more energy-efficient buildings and forms of transport, and allows broader sharing of scarce urban amenities. Other social scientists and urban planners, along with many policymakers, share this fondness for density and would like to see it increase in cities everywhere, including the densest ones.

We share some of that enthusiasm, but we also recognize that high density is synonymous with crowding. Indeed, there is a meaningful trade-off between the benefits and costs of density, and it is not clear that these benefits and costs are appropriately weighted by either market or political forces. One reason for this is that the benefit-cost calculation looks very different for insiders, long settled in the city, compared with outsiders considering moving in. In addition, the benefits and costs often operate at very different spatial and temporal scales, so they are not necessarily incorporated by all city constituents.

Understanding density is also tricky because density is both a cause and a consequence of the evolution of cities. Anything that makes a city relatively more attractive (such as a productivity increase or improved amenities) draws population from other places, which puts upward pressure on house prices, which in turn translates

■ *Gilles Duranton is Dean's Chair in Real Estate Professor at Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania. His email address is duranton@wharton.upenn.edu. Diego Puga is Professor of Economics at Cemfi (Center for Monetary and Financial Studies), Madrid, Spain. His email address is diego.puga@cemfi.es.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.3.

into higher land prices. Faced with a higher price per unit area of land, developers opt to build with a greater capital-to-land ratio (essentially, taller buildings). Faced with a higher price per unit area of floorspace, residents opt for smaller residences. With people living on smaller dwellings in taller buildings, density increases. In this sense, density is a consequence of urban evolutions.

At the same time, density is also a cause of many significant changes happening in cities. On the production side, agglomeration economies make firms and workers more productive in dense urban environments than in other locations. The benefits of density for innovation through spillovers are harder to measure but also deemed substantial. On the consumption side, higher density brings many goods and services closer, lowering travel needs. Changes in the amount and form of transport and more energy-efficient construction allow density to mitigate total pollution, albeit concentrating exposure to it. Historically, greater exposure to pollution and disease have been some of the greatest hazards of dense urban environments, and while they have lost our attention, they remain relevant today. These pitfalls, together with greater crowding and congestion, more costly floorspace for residents and firms, and scarcer green space, imply that density also has downsides. The combined benefits and costs of higher density also lead to changes in the composition of cities, triggering changes in the quality and variety of goods and services that are available—amenities in particular.

In this paper, we discuss what economic researchers have learned about density and what we see as the most significant gaps in this understanding. We begin by describing how economic research measures density for empirical enquiries and how this measurement is rapidly changing with increasingly detailed data. We then explore the benefits and costs of density, how the trade-off between them is resolved, and the welfare effects of how market and political forces affect density.

Measuring Density

Population or employment density is often used as a summary statistic to describe the spatial concentration of economic activity. In this context, density is commonly defined as the number of individuals per unit geographic area. Such "naive density" is easy to calculate. However, it may not appropriately reflect the density actually faced by the individual or firm at hand.

One problem is that economic units are traditionally defined as aggregates of administrative units. For example, US metropolitan areas are defined based on counties, but if a metro area includes some counties with substantial rural portions, such a calculation will understate the density experienced by most economic actors. In particular, the match between urban and county boundaries is systematically looser for younger and less dense metropolitan areas in the West. An extreme example is the metropolitan area of Flagstaff, Arizona, which includes the second-largest county in the country and expands across multiple national parks, monuments, and forests.

Data have now become available with much finer geographical detail than in the past. Traditional data from statistical agencies, which were previously aggregated into fairly large and often arbitrary administrative units, are now provided at a much finer spatial resolution. For instance, the US Census Bureau now routinely releases information for more than 200,000 "block groups" instead of 3,000 counties. Also, data such as property prices or retail locations that were hard or expensive to obtain have become more broadly accessible in many countries. A variety of new digital and pictorial trails has also become available, from cellphone data tracking the location of people to high-resolution satellite imagery or street-level photography.

These newly available data offer research opportunities but also raise three questions concerning: 1) choice of scale, 2) using a single "index" measure of density, and 3) the appropriate variable of interest for measuring density. Let's discuss these in turn.

The first issue is that choosing the appropriate scale at which to measure density is specific to the particular question being raised. Some agglomeration mechanisms rely on direct human interactions, which in turn suggest that effective density should be measured at a small spatial scale. In this symposium, Rosenthal and Strange discuss the literature about agglomeration economies from short-distance interactions. The study of urban travel may require the measurement of density within a five- to ten-kilometer radius to capture the distance within which most errands take place (Duranton and Turner 2018). In contrast, the metropolitan level may be relevant to measure broad-based agglomeration effects happening through local labor markets. The choice of scale does not stop at the level of metropolitan areas. Another thread of literature, inspired by Krugman (1991), has considered the much longer distances at which physical goods, and intermediate inputs in particular, can be traded. Given our urban focus, we leave aside the concentration of economic activity at a regional scale.

The choice of scale requires data on density and its effects to match. For example, De la Roca and Puga (2017) and Henderson, Kriticos, and Nigmatulina (forthcoming) have proposed measuring "experienced density" by counting population within a given radius around each individual. De la Roca and Puga (2017) then average this measure across individuals in each city, given that they do not observe the exact location within the city of employers in their wage regression. Such experienced density, in addition to dealing with the uneven tightness of area boundaries, better captures how close the typical individual is to other people when population is unevenly distributed. To give an example at the country level where boundaries are given, the United States has nearly nine times the population of Canada with a slightly smaller surface area, so its naive density is ten times higher. And yet, walking around cities and towns in both countries, one likely perceives similar concentrations of people nearby. Indeed, the average inhabitant in Canada has about 343,000 people living within a ten-kilometer radius, compared with about 306,000 in the United States.[1]

[1] We calculate experienced density using 2010 gridded population data at 3 arc-second resolution from WorldPop (2013). We first measure the number of people within a ten kilometer radius of each cell

Instead of concentrating attention on the immediate neighborhood, a spatial decay factor giving more weight to closer neighbors may also be used. It is also possible to measure population density for fine spatial units and then to take a population-weighted average for larger units that match the dependent variable. Ciccone and Hall (1996) provide an early example of this approach. Their productivity measure is at the state level, but they compute employment density at the county level before taking an employment-weighted average by state. This weighting avoids distorting the calculation of density in large states like Texas where there are vast rural portions but the population is highly concentrated in a small number of counties.

It is tempting for researchers to define the appropriate measure scale or density as the one that yields the largest or most statistically significant coefficient in the regression of interest, either implicitly or explicitly in a horse race across different measures. This temptation should be resisted. The largest or most significant coefficient may also be the one suffering from the worst identification problems.

The second problem is that any standard density measure tries to summarize a two-dimensional distribution (individuals within an area) with a single index number. However, other "shapes" of density may matter, and alternative characteristics of cities beyond just their population and land area can now be measured at a reasonable data cost. Such characteristics include the number of centers and subcenters, the mixture of land use, the compactness of development, and more. In a study of cities in India, Harari (forthcoming) find that such variables may affect a wide range of urban outcomes.

The third consideration involves choosing the variable of interest to use when measuring density. Following the pioneering work of Ciccone and Hall (1996), much of the literature that seeks to quantify the effects of the concentration of economic activity on productivity has focused on population and employment density—a choice driven mostly by the easy availability of these data. However, a case can be made that the density of human capital (Moretti 2004) or the density of business activity in the same sector of economic activity (Henderson 1974; Moretti 2019) might be more relevant variables. Moreover, as we discuss below, empirically separating agglomeration effects within and across sectors remains largely an open empirical question.

Along with these three challenges of scale, the appropriate index number, and the appropriate variable of interest, new sources of data on location and economic activity keep opening new possibilities for analysis of density. Traditional sources of population data typically measure population at its place of residence, which can work fine if the analysis is done at the metropolitan level and most people live and

in the population grid. We then compute, for all grid cells (in the entire country in this example, or in each city when we consider US metropolitan areas below), the population-weighted average of this count of people within ten kilometers. Weighting by population is important, since otherwise we would be calculating population within ten kilometers of the average place instead of within ten kilometers of the average person.

work in the same area. However, once we start trying to measure the effects of the concentration of economic activity at a fine spatial scale, a nighttime measure of density based on residences may not match well with a daytime measure of density based on employment.

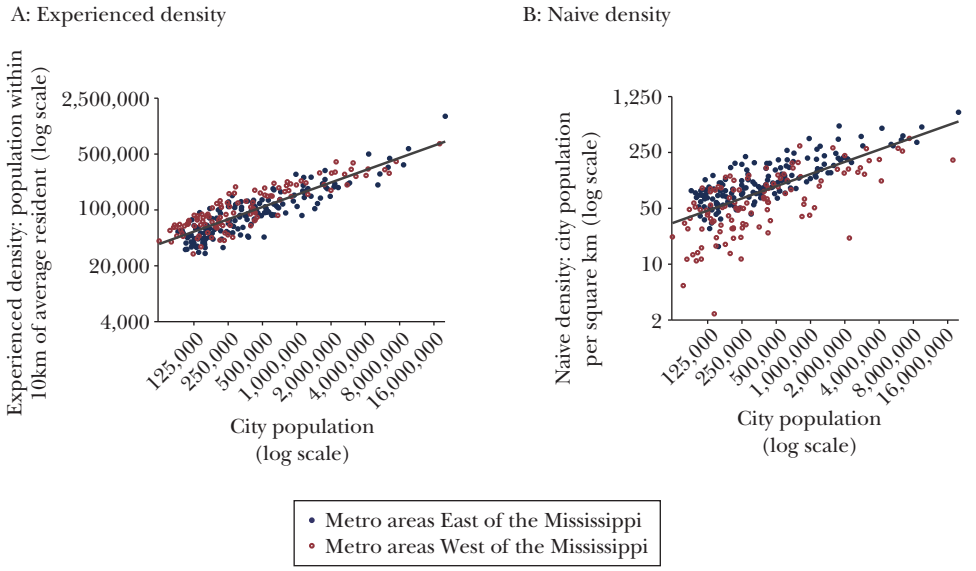
Cellphone data open the possibility of tracking people and measuring their location throughout the day (Kreindler and Miyauchi 2019). Indeed, cellphone data even allow researchers to track those interactions directly, either by studying who is talking with whom on the phone (Büchel and Ehrlich 2016; Büchel et al. 2019) or by studying who is in the same building with whom at the same time (Atkin, Chen, and Popov 2019).

Building-level data represent both day- and night-time density, and thus may offer a compromise. Daytime satellite imagery and, in some countries, official sources such as a land registry can provide detailed location data for individual buildings (Baragwanath-Vogel et al. forthcoming; de Bellefon et al. forthcoming). Information about built-up land is widely used to apportion population data measured at a broader scale to produce finely "gridded" population data (Leyk et al. 2019). "Night-lights" satellite imagery offers the possibility of easier comparisons across countries and does not rely on the availability of more traditional administrative sources. However, it suffers from a range of measurement problems, notably, the glow from bright sources of light—as discussed in this journal by Donaldson and Storeygard (2016). Building-level data, combined with population data, have the added benefit of being able to distinguish between population per unit of land area and population per unit of floorspace, which measures crowding more directly.

We have seen that most density measures either count individuals for comparable units or normalize this count by the geographical size of each unit. This raises a long-standing question: Should we measure the concentration of economic activity with its overall scale or its density? Induction suggests that, taken to extremes, neither density alone nor scale alone are particularly appealing. For instance, a highly concentrated but tiny cluster of economic activity is unlikely to generate strong agglomeration economies. On the other hand, workers located at the edge of large metropolitan areas are unlikely to benefit from their full scale in the job-matching process. The theoretical literature is mostly agnostic in the density versus scale debate. While the bulk of the work modeling the microfoundations of agglomeration economies focuses on scale effects (Duranton and Puga 2004), this is mostly a modeling choice, and it is easy to model agglomeration effects stemming from local density (Ciccone and Hall 1996).

Empirically, the relationship between city density and city population is very tight, provided we measure density carefully enough. Panel A of Figure 1 plots for US metropolitan areas experienced density, measured as population within ten kilometers of the average resident, against total population. The implied elasticity is 0.51. If we use, instead, "naive density," dividing total population by total land area within the official boundaries of the metropolitan areas, we find the same elasticity with respect to total population, 0.51, but the fit is poorer with an R^2 of 0.49 instead

Figure 1
Density versus Population for US Metropolitan Areas



Source: Authors' calculations based on data from WorldPop (2013) and US Census Bureau (2011). See footnote 2 for details.

of 0.76. This poorer fit is evident in Panel B, which also shows this is mostly because of artificially low densities in metropolitan areas with large rural portions in the western United States.[2]

The Benefits from Density

Productivity Benefits

Quantifying the productivity benefits from density has been a core theme in urban economics for several decades, and there is now broad consensus on their magnitude. The meta-analysis of Ahlfeldt and Pietrostefani (2019) suggests an elasticity of productivity with respect to density of 0.04 based on a citation-weighted

[2]We calculate experienced density using 2010 gridded population data at 3 arc-second resolution from WorldPop (2013) as detailed in footnote 1 above. We calculate naive density using population and land area data from the 2010 US Census (US Census Bureau, 2011). To define cities, we use Metropolitan Statistical Area (MSA) and Consolidated Metropolitan Statistical Area (CMSA) definitions outside of New England and New England County Metropolitan Area (NECMA) definitions in New England, as set by the Office of Management and Budget on June 30, 1999. This defines 275 metropolitan areas in the conterminous United States.

average of estimates in the literature, on Combes, Duranton, Gobillon, Puga, and Roux (2012), and on an earlier meta-analysis (Melo, Graham, and Noland 2009). Because the research on this topic has been reviewed carefully and extensively elsewhere (Rosenthal and Strange 2004; Puga 2010; Moretti 2011; Combes and Gobillon 2015), we focus this section on how such estimates are done and some recent developments.

Most estimates of the productive benefits from density are obtained by comparing productivity or earnings across spatial units with different densities. Early studies of productivity, starting with Sveikauskas (1975), studied average output per worker in cities. More recent studies rely on total factor productivity estimated from plant-level data to account for systematic differences in factor usage. In the case of earnings, firms must compare the wages they pay to the productivity benefits they receive when choosing a location. Both productivity and earnings are systematically higher in denser cities.

A concern when regressing productivity on density is that higher productivity in denser areas does not necessarily reflect a causal relationship. Instead, perhaps firms and workers are attracted to places with a strong but unobserved productivity advantage. Four strategies have been used to tackle this potential omitted variable problem. All of these approaches suggest that, while productivity-based sorting is a relevant concern, there is indeed a causal relationship in which greater urban density leads to higher productivity.

The first strategy uses instrumental variables when estimating the current density of an area. The usual instruments are historical measures of density (Ciccone and Hall 1996) and land fertility (Combes et al. 2010). Both rely on differences in density being persistent over long periods, while the determinants of productivity have changed dramatically as the economy has evolved from being mostly agricultural to a concentration on manufacturing and services. Another common instrument is land suitability for the construction of tall buildings (Rosenthal and Strange 2008; Combes et al. 2010). A limitation of these approaches is that past populations and the nature of soils may affect current productivity through the persistence of productive infrastructure or the ease of building it.

A second strategy is to include either location or plant fixed effects in an attempt to capture any unobserved attributes that may have attracted more establishments to a given city (Henderson 2003). Estimates are then identified from relating changes in productivity to changes in density over time, so the usefulness of this strategy is limited by the fact that relative changes in density tend to be small and slow (the same fact that makes the usual instruments relevant).

The third strategy is to find a quasi-experimental setting. For example, Greenstone, Hornbeck, and Moretti (2010) estimate changes in total factor productivity for incumbent plants in US counties that attracted a new plant investing over $1 million. When compared to changes in total factor productivity for incumbent plants in runner-up counties that were being considered as an alternative location by the firm, the firm's final choice can be seen as an exogenous increase in density.

The final strategy is to impose more theoretical structure on the problem, as in Baum-Snow and Pavan (2012) or Ahlfeldt et al. (2015). The latter build a quantitative framework based on a canonical urban model and apply it to Berlin, Germany, as the Wall was built and then torn down.

Another important identification issue is sorting. Larger and denser metropolitan areas disproportionately attract more educated workers. While one can control for education and other observable characteristics, unobservable worker traits that affect productivity may differ systematically across cities. Here, following Glaeser and Maré (2001) and Combes, Duranton, and Gobillon (2008), the usual strategy is to introduce worker fixed-effects when relating individual earnings to density. The productivity benefits of density are then identified from the changes in earnings that a given worker experiences when changing work location.

Higher unobserved ability may be intrinsic to a worker due to natural talent or upbringing, but it may also be something that develops over time as the worker accumulates job experience. Separating the intrinsic and experience components of ability helps to evaluate the importance of sorting. It also allows us to study the extent to which the productive benefits of density can be absorbed almost immediately or instead accumulate gradually (Glaeser and Maré 2001). De la Roca and Puga (2017) address this distinction by tracking the experience accumulated by workers in different locations. They then estimate an earnings regression where, in addition to incorporating worker fixed effects, they let the value of experience differ depending both on where it was acquired and where it is used. They conclude that workers across cities with different levels of density are not particularly different to start with; instead, working in different cities is mainly what makes their earnings diverge over time. They find that about one-half of the benefits of density are static and tied to currently working in a denser city. The other half accrues over time as workers accumulate more valuable experience in denser cities. Furthermore, workers take these dynamic gains with them when they relocate, which the authors interpret as evidence of important learning benefits to working in denser cities that get embedded in workers' human capital. These gains are stronger for those with higher initial ability.

Employing a similar strategy to look at the productivity of firms relative to density is difficult. Plant relocations are much less frequent than worker relocations. Which firms enter a market and which firms are able to survive may also be systematically different across more and less dense cities. Combes et al. (2012) develop a framework to distinguish between agglomeration and firm selection. The key insight of their model is that stronger selection in denser cities left-truncates the productivity distribution by removing the least productive firms. Stronger agglomeration instead right-shifts the productivity distribution by raising the productivity of all firms. If more productive firms benefit from density to a greater extent, this additionally dilates the productivity distribution. Using these insights, French establishment-level data, and a new quantile approach, Combes et al. (2012) show that firm selection cannot explain spatial productivity differences. Instead, there are productivity benefits from density that are even greater for more productive firms. Gaubert (2018) argues that if, as shown by Combes et al. (2012), more productive

firms benefit more from density, they will also sort into denser environments to start with. Her results indicate that sorting reinforces agglomeration economies in explaining spatial productivity differences.

Seeking the Sources of Productivity Benefits

While urban economists broadly agree on the magnitude of the productivity benefits of density, the evidence distinguishing between possible sources is less solid. Duranton and Puga (2004) classify the mechanisms into three broad classes. First, a larger market allows for a more efficient sharing of local infrastructure, a variety of intermediate input suppliers, or a pool of workers. Second, a larger market also allows for better matching between employers and employees, or buyers and suppliers. Finally, a larger market can also facilitate learning, by facilitating the transmission and accumulation of skills or by promoting the development and adoption of new technologies and business practices.

On the empirical side, a widely used strategy to distinguish between mechanisms is to measure the geographical concentration of different sectors and regress this on proxies for the different mechanisms (Audretsch and Feldman 1996; Rosenthal and Strange 2001). Because plants in any given industry are similar in many dimensions, Ellison, Glaeser, and Kerr (2010) suggest instead studying which similarities across industries help to predict better co-agglomeration patterns. Both strategies rely on having good proxies for the underlying sources of agglomeration, and how one measures these can have an important effect on results. For instance, Overman and Puga (2010) suggest that, when measuring the importance of buyer-supplier relationships, one cannot just look at the value of input purchases, but instead should focus on purchases of crucial inputs whose production is geographically concentrated.

Instead of running a horse race between different agglomeration mechanisms, another possibility is to try to isolate a particular one. This approach is challenging because of behavior and outcomes that are difficult to track. Consider knowledge spillovers. Each of the links—like those from density to additional interactions, from interactions to information flows, and from information flows to innovation—is very difficult to trace and measure. In what has for a long time been arguably the best empirical evidence of knowledge spillovers. Jaffe, Trajtenberg, and Henderson (1993) show that an inventor of a patent is more likely to live in the same location as an inventor of a patent it cites than to live in same the location as an inventor of a similar matched patent it does not cite. However, as ingenious as this strategy is, it infers interactions from spatial proximity and patents give only a very partial view of innovative activity. The strategy cannot show whether density increases interactions nor whether those interactions affect innovation more broadly.

One way to measure interactions is through survey data. Charlot and Duranton (2004; 2006) study the use of various communications technologies within and across firms in France. Their results suggest that dense urban environments result in more communication across firms and that greater communication results in higher wages, but they find little support for the hypothesis that density increases face-to-face interactions.

More recently, anonymized call detail records from cellphone operators allow measuring who interacts with whom on the phone. Büchel and Ehrlich (2016) use a major overhaul of public transport routes and schedules in Switzerland as a source of exogenous variation to show that proximity (measured by shorter travel times) does make interactions more likely. Interestingly, they find that, as in the model of Berliant, Reed, and Wang (2006), density facilitates meeting people, but this in turn makes people more choosy about with whom they interact. Thus, people in denser areas do not interact with more people, but those with whom they interact are better matches. In addition, social networks in dense urban environments are less characterized by clustering into relatively isolated groups, likely facilitating more widespread information flows.

The idea that density facilitates the quality more than the quantity of matches is also present in labor markets. Dauth et al. (2018) use matched employer-employee data for Germany to show that high-quality workers (those who get high wages conditional on observables) are more likely to work for high-quality firms (those who pay high wages conditional on observables) in denser cities. This assortative matching reinforces the fact that high-quality workers and firms are also more likely to locate in denser cities.

Modern cellphones can also provide information on users' locations, gathered from the identifier of the cell tower providing coverage to the user (stored by cellphone operators) or from location data collected by smartphone apps (purchased, combined, processed, and resold by several private companies). These data can measure spatial proximity of users at a fine geographical scale and within a narrow period: for example, two people spending more than 15 minutes in the same coffee place within the same clock-hour. Atkin, Chen, and Popov (2019) use such data to study how chance meetings contribute to innovation. They isolate smartphone users who work in buildings belonging to tech companies in Silicon Valley and trace instances where the users are in the same place at the same time. They separate chance from planned meetings and show that chance meetings result in more patent citations across firms in different sectors whose workers had met by chance more often. Other interesting new sources of data starting to be used to track specific agglomeration mechanisms include detailed input-output links between firms (Bernard, Moxnes, and Saito 2019) and search and matching in job platforms (Marinescu and Rathelot 2018).

Accessibility Benefits from Density

All else equal, having the same population of residents and establishments in a smaller area will reduce bilateral distances. However, shorter bilateral distances may encourage more trips, and more trips within a compact area may also make travel slower. What is the net effect of these influences? Using US travel survey data, Duranton and Turner (2018) estimate an elasticity of the distance traveled by an individual driver with respect to the density of workers and residents within a ten-kilometer radius around the driver's residence of –0.13, which occurs despite a very small increase in the number of trips by this driver. Travel speed also declines

with density with elasticity −0.11, but because of reduced distances, total time spent traveling declines. After considering many alternatives, Duranton and Turner (2018) conclude that the density of resident population and employment within a five- or ten-kilometer radius is the main local characteristic explaining distances traveled by local residents. Looking at credit card records for shopping and the purchase of personal services, Agarwal, Jensen, and Monte (2019) also document a decline in travel associated with a greater density of sales locations. Couture (2016) finds similar results when focusing more narrowly on restaurants.[3]

But the accessibility benefits from density cannot be captured by transport costs alone. The variety, prices, and quality of available goods and venues will all change with density. In turn, these changes affect the choices made by consumers. Regarding prices and quality, Handbury and Weinstein (2015) find that larger (and thus denser) cities do not have significantly different prices for the exact same grocery products. If prices for a certain type of product tend to be higher in large metropolitan areas, it is because consumers tend to buy higher quality varieties of the same product—like organic instead of regular eggs.

Regarding variety, Handbury and Weinstein (2015) find that the availability of grocery products, measured at the bar-code level, is much greater in larger cities. The count of restaurants accessible within a given travel time also increases with density. To assess the benefits of this expanded variety, Couture (2016) estimates the elasticity of substitution between restaurants, where a lower elasticity implies a greater willingness to pass many restaurants to access one's preferred choice. Couture's estimate of about nine for this elasticity is larger than the usual estimates for consumer goods but low enough to generate frequent trips beyond the closest restaurant and substantial welfare gains from restaurant density.

Couture's (2016) work provides an important bridge with earlier transportation research that attempts to model accessibility within a discrete choice framework (following the influential research of Ben-Akiva and Lerman 1985). After making some distributional assumptions about the preference parameters, it becomes possible in this framework to recover accessibility for a given location from the consumers' choice set of destinations in this location and the costs of reaching these destinations. For many years, the application of this approach was limited by the paucity of data about the possible destinations and the cost of reaching them. New data from sources like Google Place and Google Maps have eased these constraints. However, one limitation of these accessibility measures is that they take as given both the location of the origin of trips and the set of destinations, whereas density matters partly because it changes the set of potential venues and, as a result, possibly alters the choice of residential location.

[3] While household travel for consumption is "local," it is not "extremely local." Using Yelp data for restaurant visits in New York City, Davis et al. (2019) find that consumption is, by their metrics, about half as segregated as residences. Although transport frictions matter, they nonetheless find that social frictions play a bigger role.

Another strand of the research literature provides full general equilibrium models which consider an explicit geography and can be quantified to estimate policy counterfactuals. Redding and Rossi-Hansberg (2017) provide an excellent guide to this literature. Among models that consider urban space and transport, Ahlfeldt et al. (2015) is a particularly accomplished contribution. They model the development of a city where residents choose their residence and workplace locations and use this to explore the benefits of density using historical variation in accessibility due to the Berlin Wall. These location choices are, to a large extent, guided by utility shocks over particular commuting routes. While this approach greatly simplifies the derivation of their model, it is limiting for current-day applications to the extent that commutes represent less than one-fifth of all trips and about one-fourth of the mileage for US drivers. A challenge for the future will be to harness the recent modeling advances in economic geography, while keeping the versatility and ease of implementation of standard discrete choice approaches used in the transport literature and also making use of the much richer data now available to study urban travel.

Although we have discussed the productivity and accessibility benefits of density separately, they are interrelated. For example, the better accessibility of a denser urban environment may allow workers to search for better labor market opportunities (Manning and Petrongolo 2017). In a study of relocating research and development establishments, Xiao and Wu (2020) find that researchers who end up with longer commutes to their workplace see a drop in patenting activity while those who get closer become more productive.[4]

One possible way to integrate the productivity benefits from agglomeration into a transport and accessibility framework is to compute density in a location as the sum of nearby employment discounted by the travel cost of accessing it. This follows the suggestion of Graham (2007) and is related to the gravity specifications of the spatial quantitative models reviewed by Redding and Rossi-Hansberg (2017). However, it is important to consider such density elasticities with care. For example, taking an elasticity of earnings with respect to density estimated from cross-city variation and then applying it to a change in "effective density" resulting from some expansion of transport infrastructure in one city may overestimate the actual gains from the project—for instance, if we are considering a new subway line while agglomeration benefits arise from input-output relations between firms unaffected by this line.

Other Benefits from Density: Innovation, Reduced Pollution, Amenities

For brevity, we limit our discussion here to three especially important benefits of density that seem to us ripe for additional study: innovation, reduced pollution,

[4] In a very different context, Koster, Pasidis, and van Ommeren (2019) provide evidence about shopping externalities mediated by foot traffic. These shopping externalities are arguably about accessibility since they arise from transport savings for customers when visiting multiple stores, but they end up affecting the productivity of stores as reflected in the rents they are willing to pay.

and access to amenities. For a literature review that includes other benefits, see Ahlfeldt and Pietrostefani (2019).

The extreme concentration of innovative activity is reviewed in Carlino and Kerr (2015). Moretti (2019) estimates an elasticity of the number of patents per innovator with respect to the number of innovators in the same city and field of innovation of about 0.07. This estimate is arguably a lower bound: for example, it ignores the effect of the concentration of innovators in the same field on the probability of innovating, and it ignores spillovers arising from other fields of research. In a prior paper, Carlino, Chatterjee, and Hunt (2007) find a large elasticity of patenting with respect to urban density of about 0.20, reflecting both the higher productivity of research in denser places and the concentration of research inputs in these areas.

The link between density and pollution is also of particular importance. Residents in denser cities emit less greenhouse gas and fewer particulates than less dense cities (Glaeser and Kahn 2010). This result is only in part due to transport. There are large differences in emissions related to home cooling, even after conditioning out climatic differences. However, we need to know how much of the lower energy consumption in denser places is a consequence of smaller dwellings or if there is something uniquely energy-efficient about greater density. At the same time, as Carozzi and Roth (2019) note, a higher concentration of population within a city may result in greater overall exposure to pollution, even with lower emissions per person. After instrumenting for urban density, they find an elasticity of exposure to particulates (2.5 micrometers or smaller) with respect to density of 0.13 for the United States.

The presence of consumption amenities in dense urban areas influences how one perceives the rising inequality of wages. The increased concentration of educated workers in a small number of increasingly attractive cities is a salient feature of the US urban geography and arguably of many other developed countries (Berry and Glaeser 2005). If living in a dense area offers mainly negative "amenities," like crime, then the increased concentration of skilled workers in increasingly expensive cities implies that inequality is less than wages suggest. If living in a dense area offers positive amenities, then inequality will exceed what wages suggest (Moretti 2013). Diamond (2016) argues while rising skill premiums started the process of educated workers concentrating in dense urban settings, their presence then generated additional endogenous amenities, which she argues are central to reconciling observed changes in wages, rents, and the skill composition of residents across cities in the United States between 1980 and 2000. In highly granular empirical work, Couture and Handbury (2019) provide direct evidence about the importance of local amenities to explain the return of young educated workers to higher density residential areas in American cities. In turn, the increased concentration of educated workers appears to foster the development of new local amenities. This recent work is in tension with more traditional estimations of the relationship between amenities and city size building on Roback (1982), which suggest only a weak relationship between city population and amenities (Albouy 2008). Better knowledge about the formation of amenities in cities is undoubtedly a priority.

The Costs of Density

Land Prices, Housing Prices, Transport Costs, Congestion

Theory has long hypothesized that as population and density increase in a city, its benefits initially accumulate faster, but eventually, its costs dominate (Henderson 1974). Fujita and Thisse (2013) call this the "fundamental trade-off of spatial economics," because it explains both the existence of cities and their finite sizes. However, compared to research on benefits of density, there is a paucity of research on its costs, which Glaeser (2011) dubbed the "demons of density."

As a starting point, density brings an increase in land prices. For French urban areas, Combes, Duranton, and Gobillon (2019) estimate an elasticity of land prices at the city center with respect to their population of about 0.30. These comparisons across cities are made for a central location to make sure we compare likes with likes. In and of itself, a higher price for land does not represent a cost for society, but more expensive land elicits various responses. Some of these responses do create social costs, such as the use of more expensive building technologies to build higher or longer and slower trips as residents move further out and roads get more congested. Let's explore these costs in turn.

More costly land provides incentives to build taller. Ahlfeldt and McMillen (2018) estimate an elasticity of building height with respect to land prices of 0.30 for residential buildings and 0.45 for commercial buildings in the city of Chicago circa 2000. Interestingly, this elasticity about doubled over 100 years as technology made it easier to respond to high land costs by building taller. They note, however, that the elasticity of built-up floorspace with respect to land prices is only about one-third of the elasticity of building height because taller buildings are often surrounded by less tall buildings, open space, and roadway.

While taller buildings do provide more floorspace per unit of land, the marginal cost of floorspace increases with building height. Ahlfeldt and McMillen (2018) estimate elasticities of building cost per unit of floorspace with respect to building height ranging from 0.25 for small buildings to well above unity for skyscrapers. Tall buildings are not only costly to build; they also generate a range of recurring costs, including direct costs to their users. For example, Liu, Rosenthal, and Strange (2018) report that a typical tenant in a high-rise spends 23 minutes a day waiting for or riding elevators—about the same time as a typical one-way commute to work.[5]

A higher built-up density alleviates, but does not eliminate, the pressure created by higher land prices on the price of residences and offices. For French urban areas, Combes, Duranton, and Gobillon (2019) estimate an elasticity of housing prices at the city center with respect to their population of 0.11, compared with the

[5]Liu, Rosenthal, and Strange (2018) also report some countervailing benefits. For instance, the top floors in tall buildings command higher rents than all but the street-level, suggesting that poorer accessibility relative to lower floors is more than offset by better views and perhaps more prestige.

aforementioned 0.30 for land.[6] For US metropolitan areas, Duranton and Puga (2019) estimate a slightly lower elasticity of housing rents at the center of 0.07. Overall, higher demand for land at particularly desirable locations leads to an increase in floorspace density, higher prices for land and floorspace, and lower consumption of floorspace per person. These forces push towards an increase in human density per unit of land. Earlier, we provided an estimate of the elasticity of density with respect to population for US metropolitan areas of 0.51. In addition to lowering their housing consumption, residents also react to higher housing prices by moving to cheaper, less-accessible locations. When we estimate the elasticity of average distance to the center with respect to city population, we get 0.30.[7]

The trade-off between housing costs and transport costs has been at the heart of land-use modeling since the pioneering work of Alonso (1964), Muth (1969), and Mills (1967). However, this early work used a monocentric model of cities, which both captures many essential features of actual cities and also has important shortcomings. Most notably, residents in the basic monocentric model only travel to commute to their job, and they all work at the center. However, because not all travel is travel to work and not all commutes reach the center of cities, average travel increases with a city's population by far less than predicted by the monocentric model. Using transport data for US metropolitan areas, Duranton and Puga (2019) estimate that the elasticity of vehicle kilometers traveled with respect to the distance to the city of a resident household is only about 0.07. However, one key property of the monocentric model continues to hold. As households consider residences further away from the center, the lower price of the housing should be just offset by higher transport cost. Indeed, Duranton and Puga (2019) find that, just as predicted by the model, the elasticity of housing prices with respect to distance to the center is exactly the same as the elasticity of transport costs with respect to distance to the center, but with an opposite sign.

This literature suggests that cities that allow their urban fringe to expand may have more success in containing urban costs. Combes, Duranton, and Gobillon (2019) estimate the elasticity of land and housing prices at the center of French metropolitan areas with respect to either their density or their population. For housing prices, they estimate a density elasticity of 0.21 and a population elasticity of 0.11. For land prices, the density elasticity is 0.60 and the population elasticity 0.30. Since an increase in density is essentially an increase in population keeping

[6] In their model, Combes, Duranton, and Gobillon (2019) show that the ratio of the land price elasticity to the housing price elasticity should be equal to the share of land in construction. For France, this ratio of 0.11/0.30=0.37 is very close to the share of land in the construction of single-family homes.

[7] To estimate the elasticity of average distance to the center with respect to city population, we first determine the location of the center of each metropolitan area from the location of its core municipality reported by Google Maps. We then compute for each metropolitan area, the population-weighted average distance to the center of its Census block groups using five-year 2008–2012 data from the 2012 American Community Survey obtained from the IPUMS-NHGIS project (Manson et al. 2019). We finally regress the log of average distance on the log of population across metropolitan areas using ordinary least squares.

built-up area constant, these differences indicate that if cities could only increase their population by increasing density, house prices would increase by twice as much in the long run, with even more substantial short-term price hikes.

As a city both gets denser and expands outwards, population growth also puts a strain on its infrastructure, and particularly, its transport infrastructure. Urban travel gets slower as congestion worsens. Duranton and Puga (2019) estimate an elasticity of travel speed with respect to city population of −0.04 for US metropolitan areas using travel survey information. For cities in India, Akbar et al. (2019) estimate the same elasticity using travel time data from Google Maps and obtain a similar figure of −0.05.

When discussing the benefits of density, we included some endogenous changes in amenities, such as more consumption opportunities. Other amenity changes associated with density may instead constitute a cost. For instance, Glaeser and Sacerdote (1999) estimate that the elasticity of crime with respect to population for US cities is 0.16 if one focuses only on reported crime and 0.24 when one takes into account greater crime underreporting in larger cities. They find that, while the higher prevalence of crime-prone individuals in large cities plays an important role, almost as important is the fact that higher urban density makes finding a victim for opportunistic crimes easier and catching criminals more difficult. However, it is intriguing to note that in Europe, larger cities tend to suffer less crime (Ahlfeldt and Pietrostefani 2019).

We discussed earlier that density can also increase exposure to pollution from particulates, negatively affecting health. Historically, high density was also synonymous with frequent premature deaths caused by the poor hygiene of cities and the ease at which epidemics would propagate. Bairoch (1988) reports that early in the nineteenth century, rural youth were expected to live eight to twelve years longer than urban youth. In Europe and North America, urban life expectancy only overtook rural life expectancy after 1930. Urban planners tried to alleviate the burden of disease in cities not only by investing in water and sewage systems but also by building wider avenues and large urban parks and introducing regulations that limited overcrowding and improved air circulation and access to natural light (Colomina 2019). If cities are not denser today, it is partly a consequence of past diseases. And yet, the lack of social distancing that cities promote—and which gives them so many advantages—also makes them more vulnerable to pandemics even today.

While the literature on urban costs remains limited, it offers three tentative conclusions. First, the various elasticities reported here provide support for a hill-shaped relationship between the net benefits of cities and their population scale and density. Second, the top of this hill is fairly flat, so that the costs of being moderately undersized or oversized are small. Finally, the downward-sloping part of the net benefits may eventually fall steeply and more so if cities cannot adjust at both the intensive (densification) and extensive (outwards expansion) margins.

Aggregating the Costs of Density and Population

Quantifying urban costs, in all their forms, is complicated. As one example, the multiple components of commuting costs are hard to observe and even harder to

value (Small 2012). Housing costs, despite being transfers from users to owners, are also expected to capitalize many other costs. Moreover, housing and transport costs vary across locations in a city.

To assess overall urban costs, the literature has developed three approaches. A first strategy uses a standard urban model that also includes agglomeration benefits. For example, Au and Henderson (2006) solve such a model to obtain an expression for average value added in a city as a function of its population. They also estimate the relationship between value added per worker and city population for Chinese cities during a period in which migration was greatly restricted and conclude that many of these cities were grossly undersized. The great advantage of this approach is that it requires little data—essentially just population and value added. However, it also has several drawbacks. The key fundamental relationship between agglomeration benefits and urban costs is expected to be hill-shaped, and the shape of the hill will be hard to estimate unless many cities are far from their optimal size, as in China in the 1990s. Also, it is unclear which urban costs are reflected in lower value added (for example, commuting costs paid in the time of workers will be missed, while the higher market activity of transit firms in congested cities may result in higher value added).

The second approach models the choices of a consumer who needs to decide on a residential location and asks how much more costly it would be to achieve the same level of utility at the same location should the city become denser or grow in population. Using this approach, Combes, Duranton, and Gobillon (2019) assume that households have free mobility and leverage the insight that house prices will capitalize transport costs and amenities. As a result, the elasticity of urban costs with respect to city population turns out to be equal to the elasticity of house prices at the center of cities with respect to their population multiplied by the share of housing in household expenditure. As mentioned earlier, Combes, Duranton, and Gobillon (2019) also estimate that the elasticity of housing prices with respect to city density is 0.11 and fairly stable over the range of city populations observed in France. The share of income devoted to housing increases with urban population, from about 16 percent in a city with 100,000 inhabitants to 39 percent in a city like Paris. Taken together, these figures are indicative of urban cost elasticities associated with a greater density ranging from 0.03 for smaller cities to 0.08 for cities with more than ten million inhabitants. The main drawbacks of this approach are that it relies heavily on the free-mobility condition to simplify a wide array of changes associated with greater density, that it considers only monetary costs, and that it ignores endogenous amenities (whether positive or negative).

A third approach, developed by Duranton and Puga (2019), models the various costs of cities explicitly and estimates the parameters associated with these costs.[8]

[8] Desmet and Rossi-Hansberg (2013) propose a related approach with a quantified model. The model is then used to assess the effects of shutting down various forms of heterogeneity across cities rather than exploring the costs and benefits of increased density or rising population.

One advantage of this approach is that the key urban costs elasticity can be estimated based on equations of the model at three different levels of aggregation and using three different sources of variation. These approaches amount to estimating the assumed commuting cost equation (using within-city variation in travel distance across individuals), the spatial equilibrium within each city (using within-city variation in house prices across locations), and the spatial equilibrium across cities (using cross-city variation in city-center house prices). All three approaches result in a similar elasticity of urban costs with respect to city population of about 0.07. These urban costs are then further amplified by congestion with a population elasticity that they estimate at about 0.04. The main drawback relative to the previous approaches is that the modeling and data demands are even greater.

Getting Closer to Optimal Density?

The Unhappy Welfare Economics of Density

When considering the benefits and costs from density in a location, firms and workers choose based on their private benefits and costs, not on the social benefits and costs. There are two wedges between private and social that tend to push toward suboptimally low levels of density. The agglomeration wedge refers to the fact that firms and workers consider the agglomeration spillovers they may receive from others nearby, but not the agglomeration spillovers they may provide to others. Another wedge arises from the capitalization of land prices. When the land is not owned by local residents, a fraction of the net benefits from density are transferred away as rents to absentee landowners who benefit from agglomeration without contributing to it. On the other side, there is a congestion wedge pushing toward suboptimally high levels of density because the marginal cost of congestion exceeds its average.

The overall effect of these three wedges is ambiguous. We did report above that with respect to density, the congestion elasticity is estimated to be higher than the agglomeration elasticity. However, the smaller agglomeration elasticity pertains to the labor income of residents choosing a location, whereas the larger congestion elasticity pertains to their travel costs, which are much smaller than labor income. Much less is known about the wedge from land capitalization. Here, the key issue is not who owns the land, but whether agents making decisions about local density bear the full costs and benefits of such decisions. To complicate the welfare economics problem further, the development of high density over a sufficient spatial scale is subject to "all-or-nothing" decisions: That is, no firm may want to move to a newly-developed location unless other firms are expected to move as well. Large-scale development also often requires coordination among developers (Henderson 1974; Henderson and Mitra 1996), but the market for large-scale development is absent or limited in most of the world, and it remains limited in the United States. This coordination failure implies there might be too few communities and, as a result, they may be overly dense.

Putting together the near-absence of a market to provide density at scale and the various externalities associated with location decisions, it seems unlikely that the factors will precisely counterbalance each other in ways that cause market forces to provide optimal density.

The Unhappy Politics of Density

Almost everywhere in the world, land use is heavily regulated with a view to determine overall density as well as specific types of density (of people, jobs, shops, green space, and others).[9] A commonly heard criticism is that land-use policies tend to deliver suboptimally low levels of density. For example, many land-use policies aim to reduce densities through instruments such as minimum lot sizes, maximum floor-area ratios, or single-family residential zoning. Such policies are particularly prevalent in the United States, where land zoned for detached single-family homes accounts for 94 percent of all land zoned for residential use in San José, 81 percent in Seattle, 75 percent in Los Angeles, and 70 percent in Minneapolis, although only 36 percent in Washington, DC and 15 percent in New York (Badger and Bui 2019).

Many reasons have been suggested for restrictive zoning: 1) a fear by the rich that poorer residents will free-ride on public amenities (in particular, high-quality public schools) by consuming a small quantity of housing in a rich jurisdiction (Tiebout 1956; Fischel 1987); 2) a fear by risk-averse incumbents that less restrictive zoning would harm property values (Fischel 2001); 3) the possibility that costs of increasing density are more short-term and highly local, while the benefits may take more time to accrue and diffuse across the metropolitan area; 4) when there are strong preferences for particular locations, incumbent residents can act as monopolies restricting entry (Ortalo-Magné and Prat 2014; Hilber and Robert-Nicoud 2013); and 5) incumbent residents seeking to limit entry into particularly productive cities, thus maximizing their own welfare at the expense of potential newcomers (for a model, see Duranton and Puga 2019).

Overall, the main cost of overly restrictive land-use regulations for society may result from a spatial misallocation of population. Using very different models to quantify the social losses from excessive regulation, Hsieh and Moretti (2019) and Duranton and Puga (2019) both suggest that relaxing planning regulations in the three most productive US cities to the median level might generate large aggregate real gains of about 8 percent. How much of these gains can be realized would depend greatly on how rapidly urban costs increase as some cities grow well beyond their currently observed sizes. Nevertheless, these quantitative assessments strongly indicate that observed densities are far from optimal—too low in some places and too high in others.

[9] Given our focus, we do not discuss policy interventions that try to get firms or people to relocate over large distances, even if the density of origin and destination can be quite different. For a starting point to this work, see the papers on place-based policies in this issue, including Bartik on US policies and Overman and von Ehrlich on European policies. See Duranton and Venables (2018) for detailed discussions of place-based policies in developing countries.

Conclusion

Over the last ten years, the study of urban density has been revitalized by the arrival of new fine-grained data. We are increasingly able to observe key links such as face-to-face interactions for learning spillovers. Granular data about job searches and matching in cities or trades between firms are also increasingly available. Significant modeling advances have also taken place during the last decade. A new generation of general-equilibrium urban models has come of age, and their main novelty lies in their ability to handle the heterogeneity we observe in the distribution of jobs and residences. New models have been developed to distinguish between the agglomeration, selection, and sorting effects of density; to model job changes within and between cities; to provide better estimates of the costs of density; and so on. There is less to report on the front of causal identification during the last ten years. There has been a lot of empirical work around the issues surrounding urban density. However, it pushed more-or-less along the same lines as previously, with a continued emphasis on instrumental-variable estimations, the use of difference-in-difference after a plausibly exogenous shock, and the exploitation of spatial discontinuities.

Thus, as we look forward to future progress on the economics of urban density, our wish list includes novel data explorations providing a richer set of facts related to the manifestations of density, models that integrate urban mobility and consider the dynamics of buildings and construction, and rising empirical standards in the identification of causal effects.

As we read one last time the preprint version of this article while we are confined in response to the COVID-19 pandemic, the costs we incur and the benefits we receive by seeking proximity during normal times in dense urban environments have become even more prominent. The streets are free from congested vehicle traffic and the sky is unusually clear from pollution. At the same time, we miss the ideas that often arise from serendipitous encounters with our colleagues and the concentration and sanity of separate office and home environments. For many, the sudden drop in economic activity has brought much deeper troubles.

Beyond the temporary quietness, the immediate prominence of the costs and benefits of density, and the impact of the emerging economic crisis, what will be the long-run consequences of this virus for our densest cities? Pandemics have hit cities the hardest for centuries, and cities have adapted and been shaped by them—from investments in water and sewage systems to prevent cholera, to urban planning to reduce overcrowding and improve air circulation and access to sunlight in response to tuberculosis. Maybe temporary social distancing measures will also leave a permanent footprint on cities—for instance, in the form of more space for pedestrians and bicycles or a gain of outdoor versus indoor leisure environments. But the idea that this pandemic will change cities forever is likely an overstretch. Cities are full of inertia and this crisis has stressed both the costs and benefits of density. Confinement is forcing us to see both the advantages and the great limitations of online meetings relative to the more subtle and unplanned in-person exchanges. It has

made us realize that many tasks are impossible to do from home. At schools and universities, the haphazard transition to online courses may speed up their development, or it may delay it as many students have become frustrated by losing aspects of a full educational experience. For a while, some people may try to avoid dense cities for fear of contagion, but others may be drawn to them seeking work opportunities in difficult times. Perhaps one persisting lesson is that the cost of the pandemic has so far been associated more with urban inequalities than with urban density. While the consequences are hardest for lower-income households and minorities, they affect us all in profound ways.

■ *Puga gratefully acknowledges funding from the European Research Council under the European Union's Horizon 2020 Programme (ERC Advanced Grant agreement 695107—DYNURBAN) and from Spain's Ministry of Science and Innovation (grants ECO2016-80411-P and PRX19-00578), as well as the support and hospitality of the Wharton School's Department of Real Estate during his visit as Judith C. and William G. Bollinger Visiting Professor. We are grateful to Gordon Hanson, Enrico Moretti, Frédéric Robert-Nicoud, Timothy Taylor, and Heidi Williams for very helpful feedback and to Yan Hu, Claudio Lucciolletti and Giorgio Pietrabissa for research assistance.*

References

Agarwal, Sumit, J. Bradford Jensen, and Ferdinando Monte. 2019. "The Geography of Consumption." Unpublished.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel Sturm, and Nikolaus Wolf. 2015. "The Economics of Density: Evidence from the Berlin Wall." *Econometrica* 83 (6): 2127–89.

Ahlfeldt, Gabriel M., and Daniel P. McMillen. 2018. "Tall Buildings and Land Values: Height and Construction Cost Elasticities in Chicago, 1870–2010." *Review of Economics and Statistics* 100 (5): 861–75.

Ahlfeldt, Gabriel M., and Elisabetta Pietrostefani. 2019. "The Economic Effects of Density: A Synthesis." *Journal of Urban Economics* 111 (1): 93–107.

Akbar, Prottoy A., Victor Couture, Gilles Duranton, Ejaz Ghani, and Adam Storeygard. 2019. "Mobility and Congestion in Urban India." Policy Research Paper 8546.

Albouy, David. 2008. "Are Big Cities Really Bad Places to Live? Improving Quality-of-Life Estimates across Cities." NBER Working Paper 14472.

Alonso, William. 1964. *Location and Land Use; Toward a General Theory of Land Rent.* Cambridge, MA: Harvard University Press.

Atkin, David, Keith Chen, and Anton Popov. 2019. "The Returns to Face-to-Face Interactions: Knowledge Spillovers in Silicon Valley." Unpublished.

Au, Chun-Chung, and J. Vernon Henderson. 2006. "Are Chinese Cities Too Small?" *The Review of Economic Studies* 73 (3): 549–76.

Audretsch, David B., and Maryann P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86 (3): 630–40.

Badger, Emily, and Quoctrung Bui. 2019. "Cities Start to Question an American Ideal: A House with a Yard on Every Lot." *The New York Times* , June 18. https://www.nytimes.com/interactive/2019/06/18/upshot/cities-across-america-question-single-family-zoning.html.

Bairoch, Paul. 1988. *Cities and Economic Development: From the Dawn of History to the Present*. Translated by Christopher Braider. Chicago: University of Chicago Press.

Baragwanath-Vogel, Kathryn, Ran Goldblatt, Gordon Hanson, and Amit K. Khandelwal. Forthcoming. "Mixing Satellite Imagery to Define Urban Markets: An Application to India." *Journal of Urban Economics*.

Baum-Snow, Nathaniel, and Ronni Pavan. 2012. "Understanding the City Size Wage Gap." *Review of Economic Studies* 79 (1): 88–127.

de Bellefon, Marie-Pierre, Pierre-Philippe Combes, Gilles Duranton, Laurent Gobillon, and Clément Gorin. Forthcoming. "Delineating Urban Areas Using Building Density." *Journal of Urban Economics*.

Ben-Akiva, Moshe and Steven R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.

Berliant, Marcus, Robert R. Reed, III, and Ping Wang. 2006. "Knowledge Exchange, Matching, and Agglomeration." *Journal of Urban Economics* 60 (1): 69–95.

Bernard, Andrew B., Andreas Moxnes, and Yukiko U. Saito. 2019. "Production Networks, Geography, and Firm Performance." *Journal of Political Economy* 127 (2): 639–88.

Berry, Christopher R., and Edward L. Glaeser. 2005. "The Divergence of Human Capital Levels across Cities." *Papers in Regional Science* 84 (3): 407–44.

Büchel, Konstantin, and Maximilian V. Ehrlich. 2016. "Cities and the Structure of Social Interactions: Evidence from Mobile Phone Data." Univeristy of Bern Discussion Paper 1608.

Büchel, Konstantin, Maximilian V. Ehrlich, Diego Puga, and Elisabet Viladecans-Marsal. 2019. "Calling from the Outside: The role of Networks in Residential Mobility." CEPR Discussion Paper DP13615.

Carlino, Gerald A., Satyajit Chatterjee, and Robert M. Hunt. 2007. "Urban Density and the Rate of Invention." *Journal of Urban Economics* 61 (3): 389–419.

Carlino, Gerald, and William R. Kerr. 2015. "Agglomeration and Innovation." In *Handbook of Regional and Urban Economics*, volume 5A, edited by Gilles Duranton, J. Vernon Henderson, and William Strange, 349–404. Amsterdam: Elsevier.

Carozzi, Felipe, and Sefi Roth. 2019. "Dirty Density: Air Quality and the Density of American Cities." Centre for Economic Performance, Discussion Paper 1635.

Charlot, Sylvie, and Gilles Duranton. 2006. "Cities and Workplace Communication: Some Quantitative French Evidence." *Urban Studies* 43 (8): 1365–94.

Charlot, Sylvie, and Gilles Duranton. 2004. "Communication Externalities in Cities." *Journal of Urban Economics* 56 (3): 581–613.

Ciccone, Antonio, and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86 (1): 54–70.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics* 63 (2): 723–42.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2019. "The Costs of Agglomeration: House and Land Prices in French Cities." *Review of Economic Studies* 86 (4): 1556–89.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection." *Econometrica* 80 (6): 2543–94.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. "Estimating Agglomeration Economies with History, Geology, and Worker Fixed-Effects." In *Agglomeration Economics*, edited by Edward L. Glaeser, 15–65. Chicago: University of Chicago Press.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics of Agglomeration Economies." In *Handbook of Regional and Urban Economics*, volume 5b, edited by Gilles Duranton, J. Vernon Henderson, and William Strange, 247–348. Amsterdam: Elsevier.

Couture, Victor. 2016. "Valuing the consumption benefits of urban density." Unpublished.

Couture, Victor, and Jessie Handbury. 2019. "Urban revival in America." Unpublished.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Südekum. 2018. "Matching in Cities." NBER Working Paper 25227.

Davis, Donald R., Jonathan I. Dingel, Joan Monras, and Eduardo Morales. 2019. "How Segregated Is Urban Consumption?" *Journal of Political Economy* 127 (4): 1684–1738.

De la Roca, Jorge, and Diego Puga. 2017. "Learning by Working in Big Cities." *Review of Economic Studies* 84 (1): 106–42.

Desmet, Klaus, and Esteban Rossi-Hansberg. 2013. "Urban Accounting and Welfare." *American Economic Review* 103 (6): 2296–2327.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US workers' Diverging Location Choices by Skill: 1980–2000." *American Economic Review* 106 (3): 479–524.

Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in Economics." *Journal of Economic Perspectives* 30 (4): 171–98.

Duranton, Gilles, and Diego Puga. 2004. "Micro-foundations of Urban Agglomeration Economies." In *Handbook of Regional and Urban Economics*, volume 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2063–2117. Amsterdam: Elsevier.

Duranton, Gilles and Diego Puga. 2019. "Urban Growth and Its Aggregate Implications." NBER Working Paper 26591.

Duranton, Gilles and Matthew A. Turner. 2018. "Urban Form and Driving: Evidence from US Cities." *Journal of Urban Economics* 108 (1): 171–191.

Duranton, Gilles, and Anthony J. Venables. 2018. "Place-based Policies for Development." NBER Working Paper 24562.

Ellison, Glenn, Edward L. Glaeser, and William Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review* 100 (3): 1195–1213.

Fischel, William A. 1987. *The Economics of Zoning Laws: A Property Rights Approach to American Land Use Controls*. Baltimore, MD: Johns Hopkins University Press.

Fischel, William A. 2001. *The Homevoter Hypothesis*. Cambridge, MA: Harvard University Press.

Fujita, Masahisa, and Jacques-François Thisse. 2013. *Economics of Agglomeration: Cities, Industrial Location, and Globalization*. Cambridge: Cambridge University Press.

Gaubert, Cecile. 2018. "Firm Sorting and Agglomeration." *American Economic Review* 108 (11): 3117– 53.

Glaeser, Edward L. 2011. *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*. London: MacMillan.

Glaeser, Edward L., and Matthew E. Kahn. 2010. "The Greenness of Cities: Carbon Dioxide Emissions and Urban Development." *Journal of Urban Economics* 67 (3): 404–18.

Glaeser, Edward L., and David C. Maré. 2001. "Cities and Skills." *Journal of Labor Economics* 19 (2): 316–42.

Glaeser, Edward L., and Bruce Sacerdote. 1999. "Why Is There More Crime in Cities?" *Journal of Political Economy* 107 (S6): S225–S58.

Graham, Daniel J. 2007. "Agglomeration, Productivity and Transport Investment." *Journal of Transport Economics and Policy* 41 (3): 317–43.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy* 118 (3): 536–98.

Handbury, Jessie and David E. Weinstein. 2015. "Goods Prices and Availability in Cities." *Review of Economic Studies* 82 (1): 258–96.

Harari, Mariaflavia. Forthcoming. "Cities in Bad Shape: Urban Geometry in India." *American Economic Review*.

Henderson, J. Vernon. 1974. "The Sizes and Types of Cities." *American Economic Review* 64 (4): 640–56.

Henderson, J. Vernon. 2003. "Marshall's Scale Economies." *Journal of Urban Economics* 53 (1): 1–28.

Henderson, J. Vernon, Sebastian Kriticos, and Jamila Nigmatulina. Forthcoming. "Measuring Urban Economic Density." *Journal of Urban Economics*.

Henderson, J. Vernon, and Arindam Mitra. 1996. "The New Urban Landscape: Developers and Edge Cities." *Regional Science and Urban Economics* 26 (6): 613–43.

Hilber, Christian, and Frédéric Robert-Nicoud. 2013. "On the Origins of Land Use Regulations: Theory and Evidence from US Metro Areas." *Journal of Urban Economics* 75 (1): 29–43.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics* 11 (2): 1–39.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108 (3): 577–98.

Koster, Hans R.A., Ilias Pasidis, and Jos van Ommeren. 2019. "Shopping Externalities and Retail Concentration: Evidence from Dutch Shopping Streets." *Journal of Urban Economics* 114 (1): 103–94.

Kreindler, Gabriel E., and Yuhei Miyauchi. 2019. "Measuring Commuting and Economic Activity inside Cities with Cell Phone Records." Unpublished.

Krugman, Paul R. 1991. *Geography and Trade*. Cambridge, MA: MIT Press.

Leyk, Stefan, Andrea E. Gaughan, Susana B. Adamo, Alex de Sherbinin, Deborah Balk, Sergio Freire, Amy Rose, Forrest R. Stevens, Brian Blankespoor, Charlie Frye, Joshua Comenetz, Alessandro Sorichetta, Kytt MacManus, Linda Pistolesi, Marc Levy, Andrew J. Tatem, and Martino Pesaresi.

2019. "The Spatial Allocation of Population: A Review of Large-Scale Gridded Population Data Products and Their Fitness for Use." *Earth System Science Data* 11 (3): 1385–1409.

Liu, Crocker H., Stuart S. Rosenthal, and William C. Strange. 2018. "The Vertical City: Rent Gradients, Spatial Structure, and Agglomeration Economies." *Journal of Urban Economics* 106 (1): 101–22.

Manning, Alan, and Barbara Petrongolo. 2017. "How Local Are Labor Markets? Evidence from a Spatial Job Search Model." *American Economic Review* 107 (10): 2877–2907.

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles. 2019. "Integrated Public Use Microdata Series, National Historical Geographic Information System: Version 14.0." IPUMS, Minneapolis, Minnesota. http://doi.org/10.18128/D050.V14.0

Marinescu, Ioana, and Roland Rathelot. 2018. "Mismatch Unemployment and the Geography of Job Search." *American Economic Journal: Macroeconomics* 10 (3): 42–70.

Melo, Patricia C., Daniel J. Graham, and Robert B. Noland. 2009. "A Meta-Analysis of Estimates of Urban Agglomeration Economies." *Regional Science and Urban Economics* 39 (3): 332–42.

Mills, Edwin S. 1967. "An Aggregative Model of Resource Allocation in a Metropolitan Area." *American Economic Review Papers and Proceedings* 57 (2): 197–210.

Moretti, Enrico. 2004. "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data." *Journal of Econometrics* 121 (1–2): 175–212.

Moretti, Enrico. 2011. "Local Labor Markets." In *Handbook of Labor Economics*, volume 4, edited by Orley Ashenfelter and David Card, 1237–1313. Amsterdam: Elsevier.

Moretti, Enrico. 2013. "Real Wage Inequality." *American Economic Journal: Applied Economics* 5 (1): 65–103.

Moretti, Enrico. 2019. *The Effect of High-Tech Clusters on the Productivity of Top Inventors.* NBER Working Paper 26270.

Muth, Richard F. 1969. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use.* Chicago: University of Chicago Press.

Ortalo-Magné, François, and Andrea Prat. 2014. "On the Political Economy of Urban Growth: Homeownership versus Affordability." *American Economic Journal: Microeconomics* 6 (1): 154–81.

Overman, Henry G., and Diego Puga. 2010. "Labour Pooling as a Source of Agglomeration: An Empirical Investigation." In *Agglomeration Economics,* edited by Edward L. Glaeser, 133–50. Chicago: Chicago University Press.

Puga, Diego. 2010. "The Magnitude and Causes of Agglomeration Economies." *Journal of Regional Science* 50 (1): 203–19.

Redding, Stephen J., and Esteban Rossi-Hansberg. 2017. "Quantitative Spatial Economics." *Annual Review of Economics* 9 (1): 21–58.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6): 1257–78.

Rosenthal, Stuart S., and William Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies." In *Handbook of Regional and Urban Economics*, volume 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2119–71. Amsterdam: Elsevier.

Rosenthal, Stuart S., and William C. Strange. 2001. "The Determinants of Agglomeration." *Journal of Urban Economics* 50 (1): 191–229.

Rosenthal, Stuart S., and William C. Strange. 2008. "The Attenuation of Human Capital Spillovers." *Journal of Urban Economics* 64 (2): 373–89.

Small, Kenneth A. 2012. "Valuation of Travel Time." *Economics of Transportation* 1(1–2): 2–14.

Sveikauskas, Leo. 1975. "Productivity of Cities." *Quarterly Journal of Economics* 89 (3): 393–413.

US Bureau of the Census. 2011. "Intercensal Estimates of the Resident Population for Counties and States: April 1, 2000 to July 1, 2010."

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64 (5): 416–24.

WorldPop. 2013. "Global High Resolution Population Denominators Project." University of Southampton. https://www.worldpop.org.

Xiao, Hong Yu and Andy Wu. 2020. "Commuting and Innovation: Are Closer Inventors More Productive?" Unpublished.

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 27–49

How Close Is Close? The Spatial Reach of Agglomeration Economies

Stuart S. Rosenthal and William C. Strange

Cities exist because firms and workers benefit from spatial concentration. One benefit arises from the natural advantages present at some locations. Another is that spatial concentration allows for more diverse or less costly consumption by a city's residents. We will be concerned here with another force: agglomeration economies, which are production benefits that increase with spatial concentration. In considering agglomeration economies, our focus will be geographic. Implicit in the idea that spatial concentration increases productivity is another idea: the degree of proximity matters. Agglomeration economies must decay with distance. How close, then, do firms and workers need to be to each other to benefit from agglomeration economies? Or more colloquially, how close is close?

Our answer to this question draws on a range of research. Despite significant differences in data and methods, this research reaches similar conclusions. Agglomeration effects operate at various levels of spatial aggregation, including regional, metropolitan, and neighborhood scales. In fact, there is also evidence that agglomeration effects operate below the neighborhood level, including within buildings and organizations. Although agglomeration effects can extend over broad distances, they also attenuate, with nearby activity exerting the strongest effect on productivity.

■ *Stuart S. Rosenthal is Maxwell Advisory Board Professor of Economics, Syracuse University, Syracuse, New York. William C. Strange is SmartCentres Professor of Real Estate, Rotman School of Management, University of Toronto, Toronto, Canada. Their email addresses are ssrosent@maxwell.syr.edu and wstrange@rotman.utoronto.ca.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.27.

The spatial reach of agglomeration economies is important for several reasons. First, it sheds light on the forces that generate agglomeration economies (as noted in Rosenthal and Strange 2001). Marshall (1890) argues that there are three sources of agglomeration economies: input sharing, labor market pooling, and knowledge spillovers. Other microfoundations have also been proposed; for example, some of these build on Jacobs's (1961) idea that spatial concentration facilitates unplanned or random interactions (as in Vernon 1963). In considering these microfoundations, the different forces almost certainly operate at different geographic scales, implying that evidence regarding the attenuation of agglomeration economies is relevant to understanding their nature. Sharing of physical inputs, for example, is often associated with truck transport and can extend over regional distances. Labor market pooling is likely to have effects within commuting areas, which is to say at the metropolitan level. Knowledge spillovers as envisioned by Marshall (1890) are unplanned and are likely to be highly local. While it is true that information technology allows for effective communication with distant partners, these distant interactions are complementary to in-person interactions facilitated by close proximity (Charlot and Duranton 2004, 2006).

Second, the how-close question also bears on how public and private institutions affect agglomeration and their potential to increase productivity. To the extent that agglomeration economies operate at great distances, it is not possible to exclude migrant firms and workers from the benefits of agglomeration. This limits the ability of governments to internalize agglomeration economies through zoning or other mechanisms. On the other hand, if agglomeration economies were highly local, then it would be possible for a "developer" (in the sense of Henderson 1974) to control enough land to exclude, and problems associated with public goods would not be as severe. Industry parks can be seen in this sense. Similarly, the smaller the scale at which agglomeration economies operate, the greater the power of local governments—all of which have specific geographies over which they are empowered—to control agglomeration effects. Cities have the capacity to manage highly local agglomeration effects without the involvement of higher levels of government. At an even narrower level of geography, if agglomeration economies operate within individual buildings, the owners of these buildings have incentives to manage the composition of tenants through rent discounts and other devices as are often used to lure in anchor tenants.

Third, the spatial reach of agglomeration effects matters crucially for important markets, including commercial real estate and transportation, causing some locations to be valued over others. Agglomeration economies are capitalized in commercial real estate rents and prices and affect the design of transportation networks that govern the ability of workers to concentrate spatially.

Fourth, the tendency for agglomeration economies to attenuate drives urban spatial structure. This includes the existence of large metropolitan areas and industry clusters in addition to the ubiquitous downtown business district, often ringed by pockets of intensive commercial activity in suburban subcenters.

Together, these ideas suggest that the evolving nature of proximity will have implications for the future of cities. Because of the information technology revolution, distance is not the barrier it once was. In this new world, will cities retain an important role in productivity and growth? What forms and functions will future urban areas take?

In addressing these questions, this paper will review research on the attenuation of agglomeration effects, including freshly documented spatial patterns of employment that will help to motivate and guide portions of the discussion. We begin by reviewing evidence on agglomeration effects at the metropolitan level, where most prior research has focused. Our lens then narrows to the neighborhood level, and from there, below the neighborhood level, establishing that agglomeration effects not only extend across distances as broad as a metropolitan area but are also specific to neighborhoods, streets, and even individual buildings.

Agglomeration in Metropolitan Areas

How close is close? Fairly far, according to most of the approaches taken in the literature on agglomeration economies, which has largely analyzed agglomeration at the metropolitan area and regional levels.[1] Before discussing this literature, we will illustrate the patterns of agglomeration in a series of maps displayed in Figures 1 to 3. These figures show agglomeration effects that operate at high levels of spatial aggregation, as in the literature. They also suggest effects operating at a much tighter level of geography. This section will consider the former, while the latter is considered later in the paper.

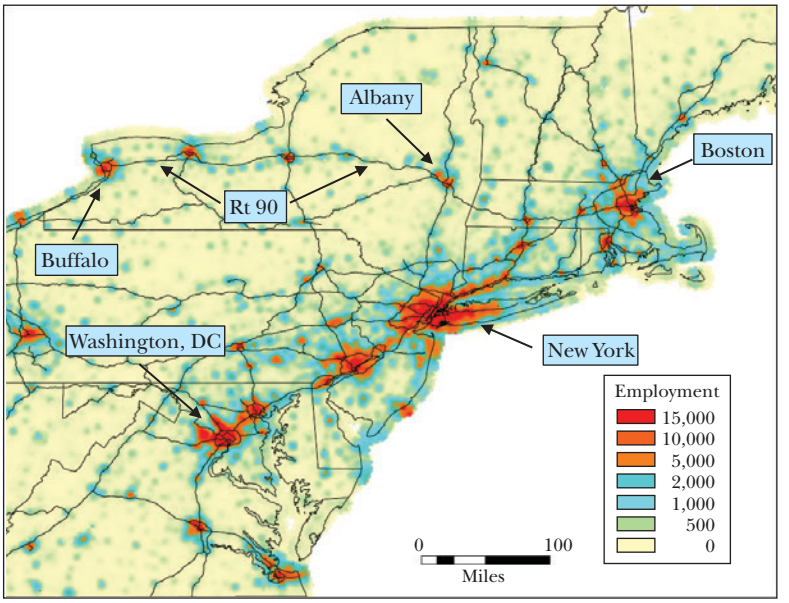
To begin, Panel A of Figure 1 displays a map of the spatial distribution of total employment across all industries for the northeast region of the United States, from Virginia and West Virginia up through northern New England. The story that the maps tell does not depend on this particular regional focus. The map was created using establishment-level data from Dun & Bradstreet for roughly 8.9 million establishments that collectively employ over 56 million workers.[2] The data were downloaded in May and June 2019 and are current as of that time. To display the data, each establishment was first geocoded at the three-meter level of precision based on its latitude and longitude reported in Dun & Bradstreet. A two-mile-by-two-mile grid was then laid down over the entire northeast region. Employment at the geographic centroid (or node) of a given grid square was set equal to the weighted sum of employment across all establishments out to 10 miles from the node. This was done using inverse distance weighting with exponential decay, so that employment

[1] On an even larger global scale, the gravity literature in international economics shows that trade between countries diminishes with distance. See Isard (1954), Isard and Peck (1954), Tinbergen (1962), and the recent survey by Head and Mayer (2014).

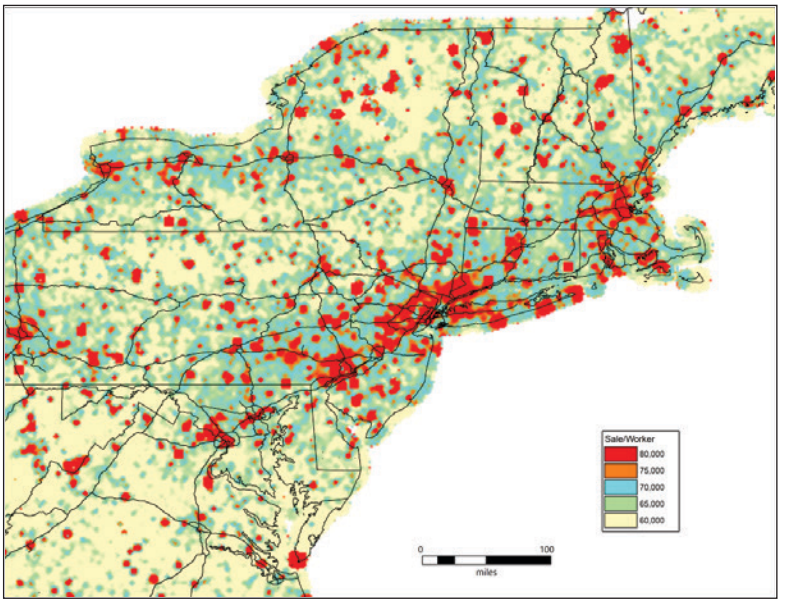
[2] Syracuse University has a site license for Dun & Bradstreet data which enabled us to work with the establishment-level information.

Figure 1
Aggregate Employment and Single-Site Average Sale/Worker Within Two Miles
(all values are smoothed out to ten miles with inverse exponential distance weighting)

A: Aggregate employment



B: Single-site sale per worker



Source: Dun & Bradstreet establishment data; US Census Tiger/Line Shapefiles.

at more distant establishments was down-weighted at an exponential rate.[3] The grid square was then assigned a color based on the level of employment assigned to its node. Calculated in this fashion, Figure 1 displays a smoothed representation of the spatial variation in employment over the northeast region.

Several patterns are apparent in the figure. First, as is well-known, employment for this region is heavily concentrated in major cities like Washington, DC, Philadelphia, New York City, Buffalo, and other urban centers. Second, concentrations of employment are also often adjacent to major interstate highways as they pass through rural areas between employment centers. This is clear along the east-west Route 90 corridor that connects Albany with Buffalo and the north-south Route 91 highway that runs up through Hartford and Springfield. Route 95 from Washington, DC, up the coast to Boston displays a nearly continuous corridor pattern of concentrated employment. These patterns contrast sharply with large areas of rural countryside that are often within a short drive of urban centers. Third, the coastal cities connected by roads comprise an industrial belt, an agglomeration of agglomerations, suggesting effects that go beyond any single metropolitan area.

The patterns in Panel A suggest that there is some aspect of the dense locations that is highly valued. Cities are expensive places to live and to do business, with high costs of labor and space. Businesses tolerate such costs only if urban locations enhance productivity by an amount sufficient to offset higher input costs. The mechanisms by which this occurs lie at the root of any answer to the question of how close a company must be to nearby activity to benefit from productivity spillovers.

Panel B of Figure 1 provides further guidance by plotting the spatial distribution of sales per worker at single-site establishments over the northeast region. Sales per worker is used here to proxy for productivity. The figure was constructed using the same Dun & Bradstreet data as above, with the sample limited to single-site firms. Restricting the sample in this fashion is necessary to ensure the accurate matching of sales to establishments.

Three patterns are especially striking relative to the employment patterns in Panel A. First, the corridor along the coast from Washington DC, up to Boston displays unusually high levels of productivity as proxied by sales per worker, mirroring employment patterns in the first panel. Second, the differences between big cities and rural areas are much smaller than in Panel A; many outlying areas also display relatively high productivity. The coast of Maine, for example, exhibits an unusual concentration of high-productivity locations. Third, the extent of variation in sales per worker across locations is far narrower than the corresponding extent of variation in spatial patterns of employment. In Panel A, there is a difference of roughly two orders of magnitude in the scale of employment between the lowest to

[3] The formula used for these purposes is given as $E_{node} = \sum_{i=1}^{n} E_i / d_i^2 / \sum_{i=1}^{n} 1/d_i^2$, where E_i is employment at establishment i located d_i miles to the grid square node, and E_{node} is the weighted sum of employment assigned to the node. For the plots in Figure 1, the search radius was set to 10 miles so that all establishments $i = 1, \ldots n$ for which $d_i \leq 10$ miles were given positive weights while establishments beyond 10 miles received zero weight. See the MapInfo manual or other standard GIS references for related details in IDW smoothing.

the highest density level indicated in the key, from below 500 to 15,000. In Panel B, the highest coded level for sale per worker (80,000) is just one-third higher than the lowest coded level (60,000).

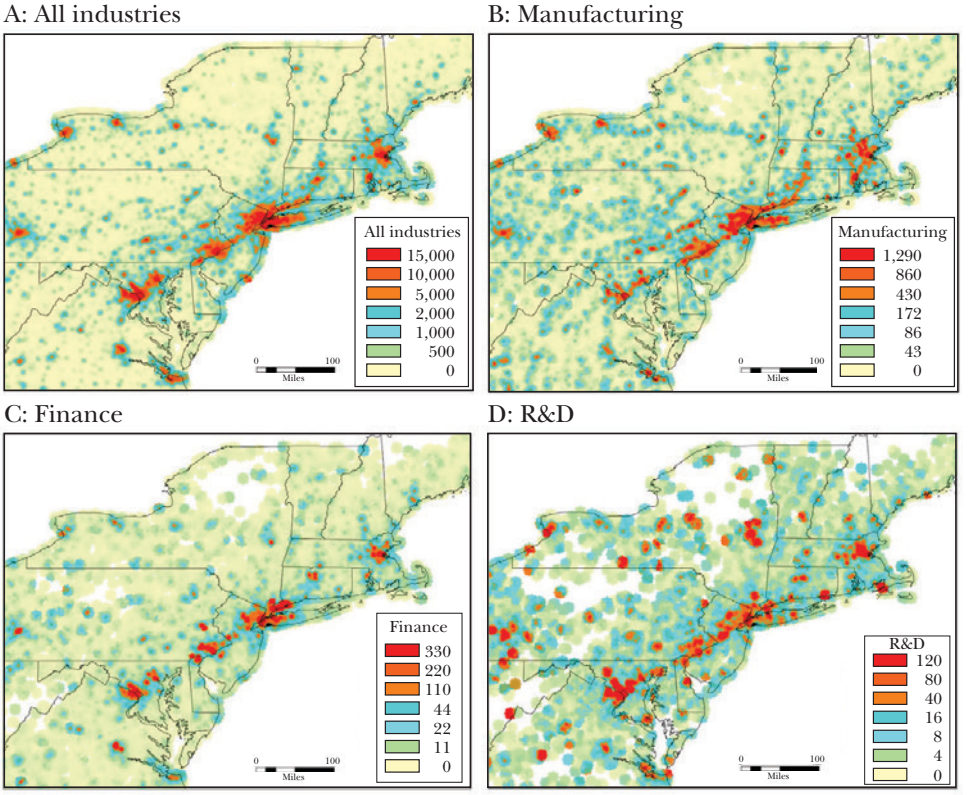
Together, the patterns in Panels A and B make clear that employment is highly spatially concentrated, while productivity, although higher in large urban centers, is much less so. This echoes evidence that doubling city size increases productivity but by a comparatively small amount—typically less than 5 percent (for example, Rosenthal and Strange 2004; Combes and Gobillon 2015; Jales, Jiang, and Rosenthal 2020). This suggests that businesses require only modest returns to choose a higher density location over a less heavily developed area. The plots in Panels A and B of Figure 1 tell a similar story.

Figure 2 revisits these issues by focusing on the clustering of industries rather than overall agglomeration. Panel A repeats the employment panel of Figure 1 for all industries combined. Panel B highlights employment in the manufacturing sector (SIC 20–39). Panel C describes employment in high value finance (SIC 62 and 67, Security & Commodity Brokers and Holding & Other Investment Offices, respectively), and Panel D plots employment in research and development (SIC 8731 and 8734, Commercial Physical & Biological Research and Testing Laboratories, respectively). To facilitate comparison of the spatial patterns across industries, the cut-off points in the keys for Panels B–D were set equal to the cut-off points used in Panel A scaled by the respective industry share of employment throughout the northeast region. Adjusted in this fashion, the relative difference in employment across the varying tone levels of shading are identical across panels. This ensures that two industries with a similar spatial distribution will have identically shaded maps. Differences across panels in employment levels for a given tone level of shading, in contrast, reflect differences in the size of the industry.

Viewed at the region level and defining industries as above, the most obvious pattern in Figure 2 is that the spatial distribution of employment for manufacturing, finance, and research and development is broadly similar to that of aggregate employment, with employment concentrated in the large cities along the corridor between Boston and Washington, DC. There are differences, however. Finance and research and development are underrepresented in rural areas, in contrast to manufacturing and total employment. Close inspection also reveals that finance is unusually concentrated in the New York metro area and that research and development is often found in localized pockets in otherwise lightly developed areas. The latter reflect in part the presence of research institutes such as the famous Woods Hole Oceanographic Institute in Woods Hole, Massachusetts, as well as research parks adjacent to rural universities as with Virginia Tech University in Blacksburg, Virginia and Cornell University in Ithaca, New York. This is consistent with research on universities as partners in knowledge creation and transmission (for example, Hall, Link, and Scott 2003; Andersson, Quigley, and Wilhelmsson 2004, 2009).

Figures 1 and 2 correspond to an extensive body of theoretical research examining agglomeration at the metropolitan level (for a survey, see Behrens and Robert-Nicoud 2015). The research builds on the theory of systems of cities

Figure 2
Employment within two Miles for Select Industries
(all values are smoothed out to ten miles with inverse exponential distance weighting)



Source: Dun & Bradstreet establishment data; US Census Tiger/Line Shapefiles.

(Henderson 1974). One conclusion of such studies is that agglomeration economies help to determine the equilibrium allocation of activity across metropolitan areas, albeit non-uniquely (Helsley-Strange 2014). Another conclusion is that agglomeration economies affect differences in factor prices across cities, including wages and rent as in Rosen (1979) and Roback (1982).

The figures are also consistent with empirical studies on the impact of agglomeration economies on spatial patterns of activity (for reviews, see Rosenthal and Strange 2004; Combes and Gobillon 2015). First, there is agglomeration of overall activity at the metropolitan level. Second, there is industry clustering (also known as "localization") at the metropolitan level.Third, there is evidence that agglomeration economies arise from Marshall's (1890) input sharing, labor pooling, and knowledge

spillovers as well as from other sources. Fourth, agglomeration economies manifest themselves in higher productivity as indicated through various measures of wages, rents, growth, and innovation.

Regarding innovation, a substantial body of evidence is consistent with knowledge spillovers at the metropolitan level. Many studies, beginning with Jaffee, Trajtenberg, and Henderson (1993), have examined spatial patterns of patent citations, while others, including Moretti (2019), have focused on patent production as an indicator of inventor productivity. Audretsch and Feldman (1996) measure new product development directly from reports of new products in industry trade journals. Andersson, Quigley, and Wilhelmsson (2009) provide evidence of knowledge spillovers by exploiting a policy-induced decentralization of higher education facilities in Sweden. Treating the establishment of new universities as exogenous, they estimate the impact of universities on indicators of local productivity and innovation. Their estimates indicate that more than half of the gain in innovative activity takes place within eight kilometers of a newly established university.[4] In complementary work, Buzard et al. (2017) show that research and development labs are spatially concentrated at various levels of geography including at the metropolitan scale, roughly. All of these studies support the conclusion that agglomeration at the metropolitan level is positively associated with innovation.

Some cautionary comments are in order. First, the papers above draw on many different data sources and methodologies, which complicates comparison across studies. Second, all studies of the impact of agglomeration on productivity must control for possible confounding effects. For example, productive workers may be drawn to large cities with attractive urban amenities, which would generate an agglomeration-productivity relationship even in the absence of agglomeration economies. Without adequate controls for such sorting, estimates may overstate the productivity gains from urbanization. For a more complete discussion of this issue, see Baum-Snow and Ferreira (2015).

A number of approaches have been taken to address these and related concerns. Obviously, richer data can help. Glaeser and Maré (2001), for example, show that the urban wage premium shrinks substantially when controls for worker attributes and worker fixed effects are included. Instrumental variable strategies have also been used, including deeply lagged regressors (Ciccone and Hall 1996) and geological variables (Rosenthal and Strange 2008a; Glaeser and Kerr 2009). Strategies based on the shape of factor return distributions have been developed in two recent papers (Combes et al. 2012; Jales, Jiang, and Rosenthal 2020). Structural approaches have also been taken (Baum-Snow and Pavan 2011), as have matching methods that exploit pseudo-natural experiments (Greenstone, Hornbeck, and Moretti 2010). Despite the very different data and approaches, all of these studies report evidence that productivity increases with city size. Moreover, recent studies

[4]Keller (2002) considers the importance of distance in international technology diffusion. See also Keller's (2004) survey.

have used increasingly rich data and powerful identification strategies, contributing to the reliability of the conclusion that agglomeration enhances productivity.

Agglomeration at the Neighborhood Level

Returning to the question of how close is close, this section will answer: close. A range of different empirical approaches find that agglomerative spillovers are stronger for agents who are closer to each other within a metropolitan area than for agents who are farther apart. This leads to the concentration of production in neighborhoods within cities, such as Wall Street.

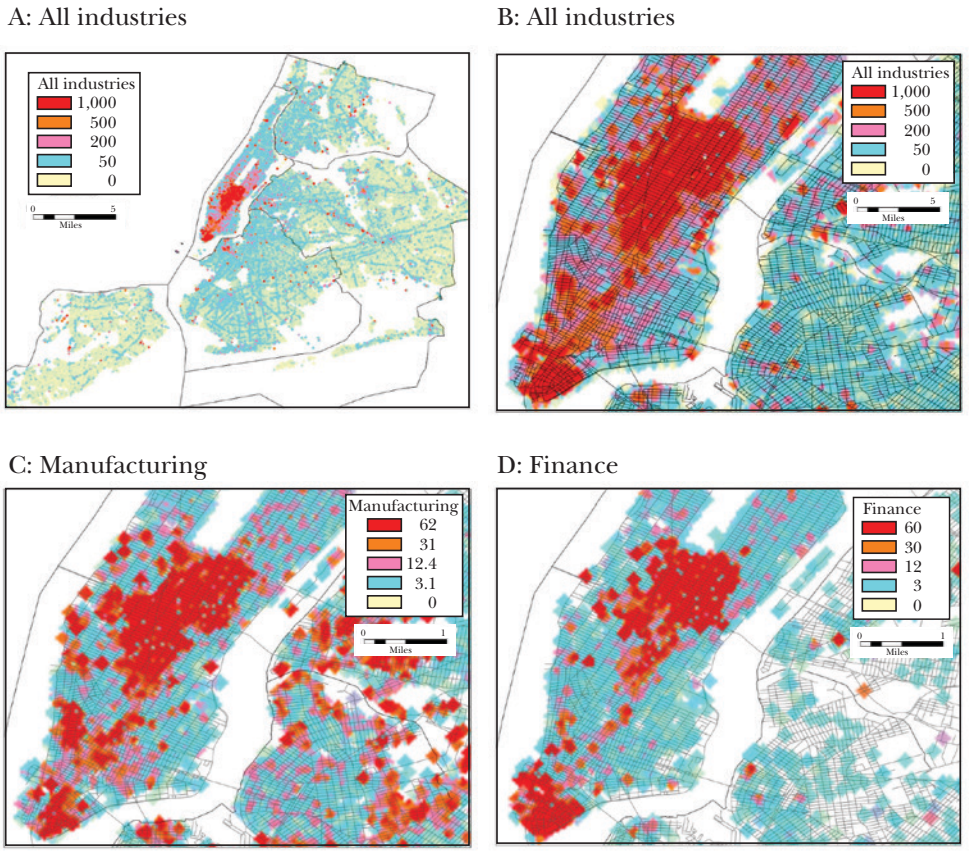
Figure 3 presents maps to illustrate this theme. Panel A displays the spatial pattern of total employment for the five boroughs (counties) that make up New York City. Panels B–D zoom in further to Manhattan and display employment patterns for total employment, manufacturing, and finance, respectively. In all four panels the data is as before, but employment is mapped at a higher level of precision, with grid squares set to 0.05 miles in width and the search radius over which employment is smoothed extending out to just 0.1 mile. For perspective, 0.05 miles is about one city block in Manhattan when traveling in a north-south direction.

In Panel A, it is apparent that employment concentration is far higher in Manhattan than in the rest of the five boroughs. Moreover, as is evident in both Panels A and B, employment in Manhattan is highly concentrated in two locations, one in Midtown, roughly between Grand Central Station and Central Park, and the other at the southern end of the island. This pattern echoes the regional pattern described above: within the largest city in the United States, employment is not uniformly distributed. Instead, it is spatially concentrated in select neighborhoods.

Panels C and D show manufacturing and finance. Once again mirroring patterns at the regional level, employment in both industries is highly spatially concentrated in select neighborhoods. This concentration takes place in different neighborhoods for the two industries and to different degrees. For manufacturing, this occurs in three zones: the area just south of Central Park, an area about halfway from Central Park to the southern tip of Manhattan, and also at the southern end of the island. For finance, employment is almost exclusively concentrated in the two dominant employment centers in Manhattan: Midtown and Lower Manhattan. The latter constitutes such a dramatic concentration of finance that it is commonly referred to as the Financial District. Outside of these areas, finance is very lightly represented and largely not present beyond Manhattan itself.

What can account for these spatial patterns? Climate obviously cannot account for spatial variation in employment density at such a narrow level of geography as in Figure 3. Proximity to port facilities matters for manufacturing but has less value to employers in finance. As a general matter, it is easier to envision a large role for amenities in explaining agglomeration at the metropolitan and regional spatial scales than at the neighborhood level. An alternative explanation is that in-person interactions between people enhance agglomerative productivity spillovers and are

Figure 3
Employment within 0.05 Miles in the Five Boroughs of New York City
(all values are smoothed out to 0.1 miles with inverse exponential distance weighting)



Source: Dun & Bradstreet establishment data; US Census Tiger/Line Shapefiles.

more prevalent for agents situated close to each other as opposed to agents who are farther apart. A growing number of studies in the literature provide support for this view, as discussed below.

The theory most relevant to understanding the patterns in Figure 3 includes Ogawa and Fujita (1980), Fujita and Ogawa (1982), and other related papers on spatial variants of agglomeration economies (for a review of this literature, see Fujita and Thisse 2013). Among the many contributions of this literature, perhaps the most important is that it solves endogenously for the location of employment instead of assuming a monocentric city. The solution depends on the tension between worker commuting costs and agglomeration economies, with the latter modeled as a spatial

spillover between firms. The former falls as employment decentralizes, with jobs located closer to where workers live. The latter rises because spillovers become weaker between firms that are further apart. The less local agglomeration economies are—in the sense of a smaller increase in the communication costs between agents as the distance between them rises—the more decentralization will be observed, both in the sense of a continuous employment gradient and in the sense of the discrete addition of subcenters. Productivity and its correlates will depend on the spatial extent of agglomeration economies in a parallel way.

There is considerable evidence that agglomeration economies attenuate, with nearby interactions having larger effects than more distant interactions. Rosenthal and Strange (2001) consider the cross-sectional pattern of localization (clustering) across industries. The paper's primary focus is the microfoundations of agglomeration economies, but the results also shed some light on attenuation. Their approach is to regress an industry's level of spatial concentration (that is, its localization) on industry characteristics. This is done at the state, county, and zip code levels of geography. Proxies for the intensity of innovative activity in an industry show a significant association with the industry's spatial concentration only at the zip code level, not at the other two levels. Proxies for input sharing, in contrast, are more strongly associated with spatial concentration at state levels. Proxies for labor pooling are significantly related to concentration at all three levels. While this does not identify the degree of attenuation of any of these three Marshallian types of agglomeration, it is consistent with knowledge spillovers attenuating the most rapidly.

Baum-Snow (forthcoming) examines the effects of highways on urban spatial structure. In addition to showing that highway construction promotes decentralization, this paper also has implications regarding attenuation based on the principle that the introduction of a highway reduces the cost of accessing central city locations (as in Baum-Snow 2007). A structural model is then used to back out a company's preference for a central city location relative to a suburban one in the same metropolitan area. The analysis suggests a large elasticity of productivity with respect to more heavily populated central city locations, implying that agglomeration effects are localized.

Business start-ups are also affected by the level and composition of nearby activity. Rosenthal and Strange (2003) work with two such measures: the number of new establishment births and the employment at these new establishments. These are separately regressed on measures of nearby activity for US data and a subset of industries, expressed as the amounts of own-industry and all-industry activity within five miles and for other distance rings beyond five miles. The marginal effect of employment in the five-to-ten mile ring is roughly half of the effect in the zero-to-five mile ring. Rosenthal and Strange (2005) carry out a parallel analysis for New York City only. This paper allows for differentiation between the effects that are within one mile and one-to-five miles away. Again, there is sharp attenuation. The within one-mile effect is roughly twice as large as the one-to-five mile effects for both establishment births and new establishment employment. In a similar vein, Arzaghi and Henderson (2008) consider New York's advertising industry, historically located

around Madison Avenue in Midtown. They estimate a (Poisson count) model of openings of new single-site advertising companies as a function of proximity to other nearby advertising agencies, along with other controls. They find evidence of significant spillovers between advertising companies, with effects that largely attenuate within roughly 750 meters. They argue that their findings are likely reflective of knowledge spillovers, in part because of the highly localized pattern of estimated effects.

Other papers have looked at productivity and its correlates. Rosenthal and Strange (2008a) estimate wage models.[5] Unlike Glaeser and Maré (2001) and most of the rest of the urban wage literature, the paper defines geographic units based on continuous distance measures rather than relying on political boundaries (as with states or counties, for example). Specifically, it examines the relationship between wage and the amount of employment within five miles and between five and twenty-five miles, controlling as usual for worker characteristics. The paper considers two sorts of local density within each distance band: the density of workers with college or university degrees and the density of workers without these degrees. Geological variables related to the cost of density—access to bedrock, seismic and landslide hazard, coverage by water—are used to instrument for the employment regressors. The effect of nearby college-educated workers is significant and positively related to wage, but the effect of more distant college-educated workers is close to zero. Concentrations of nearby non-college workers, in contrast, significantly reduce wage but also with a sharp attenuation pattern. The latter result is a reminder that agglomeration without sufficient positive spillovers can impede productivity by contributing to congestion.[6]

A very different set of papers has examined the potential for residentially based labor market networks to increase productivity by enhancing the quality of labor market matching between workers and employers. Using confidential census data, Bayer, Ross, and Topa (2008) show that workers who live within the same census block are more likely to work at establishments close to each other than individuals who live only a modest distance further apart. Using matched employer-employee data, Hellerstein, McInerney, and Neumark (2011) find a similar pattern, and Hellerstein, Kutzbach, and Neumark (2014) show that job turnover and wages vary with social connections within a residential neighborhood in ways that support the idea that increased neighborhood connectedness enhances worker productivity. Although this literature does not provide evidence on spillovers between employers, these papers further confirm the general principle that neighborhood-level proximity can foster productive interactions.

[5] Moretti (2004) documents the existence of large human capital spillovers at the metro level.

[6] Analogous attenuation patterns are also evident in Li (2014). Li shows that a greater concentration of in-state doctors within 25 miles lowers mortality rates from various diseases relative to similar concentration of more distant doctors. She also shows that state borders reduce the positive effect of nearby medical personnel, consistent with state licensing laws that restrict the ability of doctors to treat patients across state lines. This result provides evidence that local government policy can affect the transmission of agglomeration economies, in this case with negative effect.

Of course, agglomeration effects will be captured not just in wages but also in rents, as in Rosen (1979) and Roback (1982). Wage estimates therefore may capture only part of the agglomeration effect. This suggests a research strategy of studying the relationship between agglomeration and the commercial or industrial rent paid by the tenant. Unfortunately, these data are not commonly available, and the great heterogeneity of commercial and industrial real estate means that it will be difficult to gather the sort of data that allows an apples-to-apples comparison.

To overcome the difficulty of obtaining useful rent data, Liu, Rosenthal, and Strange (2018a) work with confidential offering memos that report rent. The cost of this resolution is that the data are non-representative in that offering memos are generated only when buildings are put up for sale. However, this paper obtains another result consistent with agglomeration economies operating at the neighborhood level, showing that rents are positively related to the intensity of activity within a building's zip code. The point estimate suggests that doubling employment within the zip code is associated with a roughly 11 percent increase in commercial rent. These effects are found for office industries such as law, finance, and business services—precisely the industries that have come to dominate downtowns.

Ahlfeldt et al. (2015) takes a different perspective to the attenuation of agglomeration effects by using the exogenous variation in nearby density associated with the construction and demolition of the Berlin Wall. Reduced form estimates show that the Wall hindered access to those parts of the prewar central business district located in East Berlin. Structural estimates of the attenuation parameter imply highly localized productivity spillovers, with effects reaching roughly zero at ten minutes of travel time. This corresponds to about half a mile by foot and 2.5 miles by subway (respectively, 10 and 50 Manhattan blocks). This is yet another approach, one with strong identification and tight ties to theory, that finds the same result of rapidly attenuating agglomeration effects.

Despite differences in approach, the papers above reach similar conclusions: agglomeration economies attenuate rapidly. For example, Rosenthal and Strange (2003) conclude that spillover effects shrink by roughly half after five miles, while Rosenthal and Strange (2005) find effects that are notably smaller after one mile. Arzaghi and Henderson (2008) report evidence that among advertisers, spillovers attenuate away within 750 meters, or a little less than half a mile. Although measured based on travel time and not distance, results from Ahlfeldt et al. (2015) similarly suggest rapid attenuation. There is thus a clear consensus that proximity matters.[7]

There are several reasons why the attenuation of agglomeration effects matters, as mentioned earlier in the introduction. As shown by Fujita and Ogawa's work (1980, 1982), attenuation of agglomeration economies has an important effect on urban spatial structure. It determines whether a city is monocentric and if so, how spatially concentrated employment may be. It determines whether subcenters form,

[7] It should also be noted that even if agglomeration economies were entirely local, we would still observe agglomeration at a much larger scale due to the overlap of local clusters (as noted by Kerr and Kominers 2015).

and if so, how many.[8] More generally, it determines the degree and form of urban sprawl.

The robust result that agglomeration effects are local also has implications for the microfoundations of agglomeration economies. Marshall (1890) identifies knowledge spillovers, labor pooling, and input sharing as potential sources. Jacobs (1961) emphasizes the value of unplanned synergies among residents of large cities. The results discussed above suggest that certain agglomeration forces operate when agents are close to each other. Planned and unplanned interactions that contribute to knowledge sharing are likely to be more local in nature, taking place between agents who are familiar with each other. This familiarity is likely to be tied to proximity. Labor markets tend to operate at longer distances; in fact, metropolitan areas are defined in part by commuting flows. Similarly, physical inputs are often transported great distances. This is not to say that there is not a local element to labor pooling and input sharing. Local word-of-mouth job market networks are part of labor market pooling, while input sharing sometimes involves repeated interactions that can be enhanced by face-to-face meetings facilitated by proximity (Vernon 1963). Our point is, instead, that the in-person interactions are more central to Marshallian knowledge spillovers since in-person communications are likely to be more important.

The local nature of agglomeration effects also has normative implications. Hsieh and Moretti (2019) present a quantitative model of agglomeration in order to assess the welfare consequences of land use regulation. To the extent that land use regulation is binding, it raises the cost to a city of accommodating a larger population. This, in turn, means that there is a spatial misallocation, where households and firms are not located in the places that maximize welfare. Their calibrations show a large effect. All of this analysis takes place at the metropolitan level. In this setting, the inability to develop at high density in one part of a metropolitan area (say, in very restrictive Toronto) can be overcome if another part (for instance, less restrictive Mississauga) is not similarly constrained. With localized agglomeration effects, this spatial substitution is not possible, implying that the costs associated with binding land use regulation may be even higher.

Another normative implication pertains to the role of entrepreneurial agents who profit from correcting urban resource misallocation. Henderson (1974) refers to these agents as "developers," with the idea that inefficiency will be capitalized into land prices allowing a developer to profit from welfare-enhancing policies. There are clearly no agents who can perform this role at the scale of an entire city; even the biggest developer is not this large. However, to the extent that a significant fraction of effects are localized, a developer will be more likely to be able to internalize the relevant spillovers. For instance, the developers of London's Canary

[8] McMillen and Smith (2003) estimate the relationship across a sample of cities between the number of subcenters and a city's population and commuting costs. These two variables are strongly predictive of the number of subcenters, as anticipated by the Ogawa-Fujita model discussed earlier. See also Giuliano and Small (1991) and McMillen and McDonald (1998) for further analysis of subcenters.

Wharf financial district were able to control the entire district. For further discussion of this issue, see Helsley and Strange (1997).[9]

Agglomeration below the Neighborhood Level

We now zoom in even more tightly and show that for agglomeration economies, how-close can mean very close. In addition to operating at the metropolitan and regional levels and at the neighborhood level, agglomeration economies operate well below the neighborhood level.

As one example, agglomeration economies appear to operate within individual buildings. Liu, Rosenthal, and Strange (2018b) show that office buildings are specialized even in small business districts that are themselves specialized. We provide graphic evidence of this in Figure 4. The southern end of Manhattan exhibits a well-known specialization in banking and finance (see Figure 3, Panel D). Figure 4 displays all of the buildings in this neighborhood, both in two dimensions (Panel A) and three dimensions (Panel B). In both panels, buildings with a higher finance share of employment are shaded a more vibrant tone of red. Despite the specialization of the neighborhood, most buildings actually have little or no finance, while only a relatively small number of buildings are dominated by financial services. Even within an area famous for financial services, buildings are specialized.

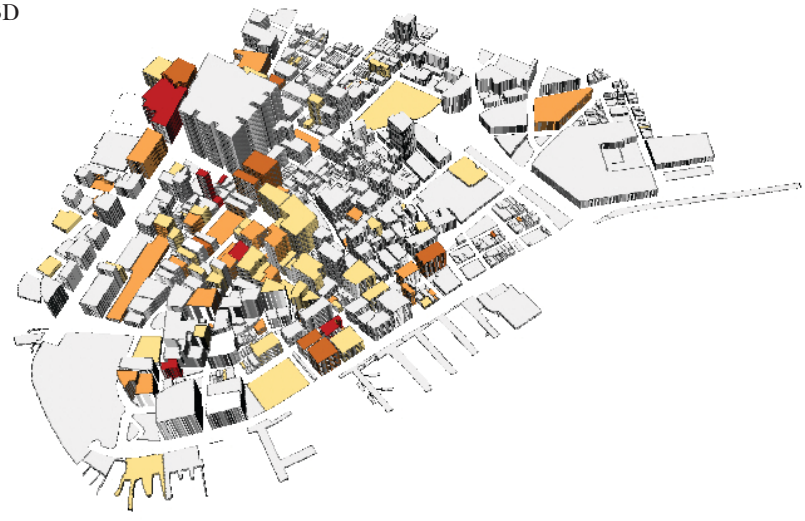
Liu, Rosenthal, and Strange (2018b) conduct a more complete assessment of these patterns for finance and other industries that dominate office buildings in city centers, such as law, advertising, and retail. For the neighborhoods adjacent to the New York Stock Exchange and Grand Central Station, commercial activity is specialized in select buildings beyond what random assignment would imply. This is true even controlling for building quality, which could potentially make some buildings better suited for specific tenants. Furthermore, for roughly 50,000 buildings in the city centers of New York, Chicago, San Francisco, Los Angeles, and Washington, DC, Liu, Rosenthal, and Strange (2018b)provide evidence that building-level productivity spillovers likely contribute to building-level specialization. The identification strategy focuses on the relationship between the presence of an anchor establishment and the composition of other commercial activity in the anchor's building and also employment in the adjacent building on the same side of the street. Controlling for building fixed effects and the composition of employment within roughly two blocks, evidence indicates that when an anchor is present, other establishments in the anchor's building display 15 to 18 percent higher employment in the anchor's own industry. This effect drops to just 1 percent, however, for the adjacent building on the same block face. These patterns support the view

[9]Another institution for internalizing spillovers is the Business Improvement District in which local business owners form an association and act as local "private governments" in order to influence the attributes of the neighborhood business environment with potential to improve efficiency (for example, Helsley and Strange 1998).

Figure 4
Finance Share of Employment (SIC 62, 67) in the Financial District

A: 2D

B: 3D



Source: Dun & Bradstreet establishment data; New York City Planning Department.

that productivity spillovers associated with proximity to anchor establishments draw complementary companies together and that such spillover effects decline sharply upon leaving the building.

In fact, specialization may take place at an even smaller geographic scale. Liu, Rosenthal, and Strange (2020) show that within tall commercial buildings, employment per square foot of office space is higher when an establishment has other establishments in its industry on its floor. This effect is also significant on the immediately adjacent floors, up or down, although it is reduced by more than half. The effect largely attenuates away by a distance of three floors. Because establishment employment density increases with productivity as a company grows and adds more workers to existing space, this pattern is consistent with within-building productivity spillovers that dissipate once vertical distance exceeds typical stairwell walking distance, at which point elevator travel is used. Thus, agglomeration effects seem to take place within buildings or even between adjacent floors in a building.

This finding leads naturally to the question of whether there are spatial effects operating even within establishments and firms. Because these effects are internal to firms, there is a sense that they are not spillovers in the classic sense. However, they are spatial effects that are external to individual workers. Charlot and Duranton (2006) document the substantial amount of communication taking place within a firm. Mas and Moretti (2009) show that the presence of an unusually productive worker in a supermarket enhances productivity of other workers in the store. This effect is strong when the productive worker is on the same shift and visible to other workers, but weak otherwise. Sandvik, Saouma, and Seegert (2019) provide analogous evidence based on an experimental design. They show that increased communication between co-workers in sales call centers increases productivity in ways indicative of knowledge sharing and learning from peers. In a completely different setting, Bosquet and Combes (2017) show that economists in French universities develop more successful publication records when there are other academics in their department with a similar field of emphasis. To the extent that spillovers are within firms and other organizations, both the capacity and the incentive to address spillovers are present.

In fact, the capacity and incentive to internalize spillovers are even stronger. We previously observed that the geographic scale of agglomeration economies had implications for the ability of agents to internalize agglomeration spillovers. While developers only rarely control entire commercial or industrial districts, individual buildings are owned by agents with the capacity and incentives to manage spillovers. This idea is familiar in the context of shopping malls, a particular type of commercial structure. In that context, it is standard practice for mall owners to seek big-box anchor tenants that are perceived as generating positive shopping spillovers that attract additional smaller tenants. This idea is found in theoretical work by Brueckner (1993) and Konishi and Sandfort (2003) and empirical studies by Pashigian and Gould (1998) and Gould, Pashigian, and Pendergast (2005). It is also present in Koster, Pasidis, and van Ommeren (2019) who provide evidence of spillovers on shopping streets outside of a mall context. The finding of highly localized

spatial interactions implies the possibility of internalization without government intervention.

All of this means that we see evidence of agglomeration effects operating at the metropolitan scale, the neighborhood scale, and below the neighborhood scale. The latter includes effects operating within individual buildings and even floors within buildings. These are very local spatial spillovers indeed.

Conclusion

How close is close? Taken together, the evidence presented in this paper shows that agglomeration effects operate at various spatial scales, with nearby effects the strongest. This pattern can reflect a number of forces. First, it may reflect a single agglomeration effect with spillovers decreasing with distance. For example, the labor pooling benefits enjoyed by employers are likely to shrink as they become farther apart, since worker commuting costs to an alternate employer will tend to increase. Second, it may reflect the combined effects of multiple agglomeration forces, where the individual forces have different ranges. Knowledge spillovers are likely to operate at a narrower spatial level than labor pooling, for example. Finally, there may be heterogeneity among agents in their interaction costs. All agents can presumably benefit from activity that is very close, but some may not be sufficiently "networked" to benefit from interactions further away. This is one way to interpret the Rosenthal and Strange (2012) analysis of female entrepreneurship, which presents evidence consistent with female entrepreneurs enjoying less benefit from agglomeration than male entrepreneurs.

The continued importance of proximity is notable in light of the huge reductions in interaction costs witnessed in recent years. In considering why proximity continues to matter, Glaeser (1998) proposes three key transport costs he sees as driving the future of cities: the costs of moving ideas, people, and goods. Road building and other transport improvements have certainly affected the costs of moving people and goods, making it easier to access employment centers from greater distance. This was documented by Baum-Snow (2007) who shows that radial urban highways contribute to decentralization of US cities and growth of the suburbs in urban areas. A parallel transport mechanism likely helps to explain the concentration of employment along major highways in otherwise rural areas, as noted earlier in the discussion of Figure 1. Analogously, Dong, Zheng, and Kahn (2020) show that the recent introduction of high-speed bullet trains in China have contributed to increased partnerships and co-authorship among scholars in universities in different cities. This reminds us that the physical transportation costs associated with interaction can also affect the cost of moving ideas, extending the spatial reach of knowledge spillovers and diffusion of ideas at the regional scale.

Since Glaeser's (1998) paper, the information technology revolution has surely affected his three sorts of transport costs in ways that at first glance might be expected to contribute to greater dispersion of activity. This includes the many

changes associated with electronic communication that have reduced the cost of sharing ideas from afar. It also includes recent innovations like ride-sharing which has reduced the cost of travel within a metropolitan area (Hall, Palsson, and Price 2018); presumably, the deployment of autonomous vehicles will reduce future travel costs.

Despite all of these innovations, we continue to see evidence of agglomeration effects operating at highly local spatial scales. These scales include neighborhoods, individual buildings, and even spatial arrangements of workers within buildings, all of which have potential to foster local interactions. It is worth noting, however, that there has been no work in the economics of agglomeration literature that has carefully considered the effect of dramatic reductions in interaction cost on changes in the spatial scale at which agglomeration economies operate. Returning to the maps from earlier in the paper, it is notable that the Northeast's large cities at the founding of the United States are mostly the large cities we see today. Of course, new cities arose in other places, but the historic cities remain important. Because the technological forces governing agglomeration have changed profoundly, this pattern implies that equilibrium patterns of agglomeration change slowly, which is consistent with evidence from Bleakley and Lin (2012) and others. Another reason for the continued importance of highly proximate interactions may be that they are complementary to more distant interactions that new technology now allows. An example would be the potential to establish partnerships in person that could then operate effectively from remote locations in subsequent years.

It is also worth emphasizing that the information technology revolution is fairly recent, and so its effects on urban form and function are likely still evolving. Online retail, for example, is new and growing rapidly. While internet purchases have the potential to draw retail activity out of city centers, online retail is not a substitute for the appeal of window shopping or the buzz of night life on a busy street. To the extent that such urban amenities have disproportionate appeal to high-productivity workers, this may contribute to gentrification and a rising concentration of college-educated residents in city centers, as recently documented by Couture and Hanbury (2019). An analogous amenity-based mechanism likely explains the tendency for high-productivity establishments to concentrate high in tall commercial buildings where views are more dramatic, as recently documented by Liu, Rosenthal, and Strange (2018). Although our focus here is on the spatial reach of productivity spillovers, localized and endogenously created amenities will contribute to concentrations of skilled workers. That, in turn, may amplify localized productivity spillovers. This would be consistent with evidence from Rosenthal and Strange (2008a, 2008b) and Mas and Moretti (2009) that proximity to productive workers tends to boost performance.

In sum, improvements in information technology have still left us with agglomeration economies that operate at both broad and narrow spatial scales. Information technology clearly allows for productive distant interactions. One example is a radiologist reading an x-ray from a remote site. Other examples include the increasing use of video conference business meetings that take advantage of increasingly

effective remote communication software, reinforced by distant interactions necessitated by the coronavirus pandemic. Nevertheless, both through direct and indirect channels, a range of evidence all points to continued benefits from proximity at narrow levels of geography, including neighborhood, building, and even within-building locations.

■ *We thank Gordon Hansen, Enrico Moretti, Timothy Taylor, Heidi Williams, Nathaniel Baum-Snow, Gilles Duranton, and Matthew Turner for helpful suggestions. We also thank Rolando Campusano, Maeve Maloney, and Joaquin Andres Urrego Garcia for valuable research assistance. Any errors are our own.*

References

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf. 2015. "The Economics of Density: Evidence from the Berlin Wall." *Econometrica* 83 (6), 2127–89.

Andersson, Roland, John M. Quigley, and Mats Wilhelmsson. 2004. "University Decentralization as Regional Policy: The Swedish Experiment." *Journal of Economic Geography* 4 (4): 371–88.

Andersson, Roland, John M. Quigley, and Mats Wilhelmsson. 2009. "Urbanization, Productivity, and Innovation: Evidence from Investment in Higher Education." *Journal of Urban Economics* 66 (1): 2–15.

Arzaghi, Mohammad, and J.Vernon Henderson. 2008. "Networking off Madison Avenue." *Review of Economic Studies* 75 (4): 1011–38.

Audretsch, David B., and Maryann P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86 (3): 630–40.

Baum-Snow, Nathaniel. 2007. "Did Highways Cause Suburbanization?" *The Quarterly Journal of Economics* 122 (2): 775–805.

Baum-Snow, Nathaniel. Forthcoming. "Urban Transport Expansions and Changes in the Spatial Structure of US Cities: Implications for Productivity and Welfare." *Review of Economics and Statistics.*

Baum-Snow, Nathaniel, and Ronni Pavan. 2011. "Understanding the City Size Wage Gap." *The Review of Eonomic Studies* 79 (1): 88–127.

Baum-Snow, Nathaniel, and Fernando Ferreira. 2015. "Causal Inference in Urban and Regional Economics." In *Handbook of Urban and Regional Economics, Volume 5A*, edited by Gilles Duranton, J. Vernon Henderson, and William Strange, 3–68. Amsterdam: Elsevier.

Bayer, Patrick, Stephen L. Ross, and Giorgio Topa. 2008. "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes." *Journal of Political Economy* 116 (6): 1150–96.

Behrens, K., and F. Robert-Nicoud. 2015. "Agglomeration Theory with Heterogeneous Agents." In *Handbook in Regional and Urban Economics, Volume 5A*, edited by Gilles Duranton, J. Vernon Henderson, and William Strange, 169–247. Amsterdam: Elsevier Press.

Bleakley, Hoyt, and Jeffrey Lin. 2012. "Portage and Path Dependence." *The Quarterly Journal of Economics* 127 (2): 587–644.

Bosquet, Clément, and Pierre-Philippe Combes. 2017. "Sorting and Agglomeration Economies in French Economics Departments." *Journal of Urban Economics*, 101: 27–44.

Brueckner, Jan K. 1993. "Inter-store Externalities and Space Allocation in Shopping Centers." *The Journal of Real Estate Finance and Economics* 7 (1): 5–16.

Buzard, Kristy, Gerald A. Carlino, Robert M. Hunt, Jake K. Carr, and Tony E. Smith. 2017. "The Agglomeration of American R&D Labs." *Journal of Urban Economics* 101: 14–26.

Charlot, Sylvie, and Gilles Duranton. 2006. "Cities and Workplace Communication: Some Quantitative French Evidence." *Urban Studies* 43 (8): 1365–94.

Charlot, Sylvie, and Gilles Duranton. 2004. "Communication Externalities in Cities." *Journal of Urban Economics* 56 (3): 581–613.

Ciccone, Anthony, and Robert E. Hall. 1996. "Productivity and the Density of Economic Activity." *The American Economic Review* 86 (1): 54–70.

Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics of Agglomeration Economies." In *Handbook in Regional and Urban Economics*, Volume 5, edited by G. Duranton, J. V. Henderson, and W. Strange, 247–348. Amsterdam: Elsevier Press.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga and Sébastien Roux. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection." *Econometrica* 80 (6): 2543–94.

Couture, Victor, and Jessie Handbury. 2019. "Urban Revival in America, 2000 to 2010." NBER Working Paper 24084.

Dong, Xiaofang, Siqi Zheng, and Matthew E. Kahn. 2020. "The Role of Transportation Speed in Facilitating High Skilled Teamwork across Cities." *Journal of Urban Economics* 115 (103212).

Fujita, Masahisa, and Hideaki Ogawa. 1982. "Multiple Equilibria and Structural Transition of Non-monocentric Urban Configurations." *Regional Science and Urban Economics* 12 (2): 161–96.

Fujita, Masahisa, and Jacques-François Thisse. 2013. *Economics of Agglomeration: Cities, Industrial Location, and Globalization.* 2nd ed. Cambridge: Cambridge University Press.

Giuliano, Genevieve, and Kenneth A. Small. 1991. "Subcenters in the Los Angeles Region." *Regional Science and Urban Economics* 21 (2): 163–82.

Glaeser, Edward L. 1998. "Are Cities Dying?" *Journal of Economic Perspectives* 12 (2): 139–60.

Glaeser, Edward L., and David C. Maré. 2001. "Cities and Skills." *Journal of Labor Economics* 19 (2): 316–42.

Glaeser, Edward L., and William R. Kerr. 2009. "Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?" *Journal of Economics & Management Strategy* 18 (3): 623–63.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy* 118 (3): 536–98.

Gould, Eric D., B. Peter Pashigian, and Canice J. Prendergast. 2005. "Contracts, Externalities, and Incentives in Shopping Malls." *Review of Economics and Statistics* 87 (3): 411–22.

Hall, D. Jonathan, Craig Palsson, and Joseph Price. 2018. "Is Uber a Substitute or Complement for Public Transit?" *Journal of Urban Economics* 108: 36–50.

Hall, Bronwyn H., Albert N. Link, and John T. Scott. 2003. "Universities as Research Partners." *Review of Economics and Statistics* 85 (2): 485–91.

Hellerstein, Judith K., Melissa McInerney, and David Neumark. 2011. "Neighbors and Coworkers: The Importance of Residential Labor Market Networks." *Journal of Labor Economics* 29(4): 659–95.

Head, Keith, and Thierry Mayer. 2014. "Gravity equations: Workhorse, Toolkit, and Cookbook." In *Handbook of International Economics,* Vol. 4, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 131–95. Amsterdam: Elsevier.

Hellerstein, Judith K., Mark J. Kutzbach and David Neumark. 2014. "Do Labor Market Networks Have an Important Spatial Dimension." *Journal of Urban Economics* 79: 39–58.

Helsley, Robert W., and William C. Strange. 1997. "Limited Developers." *Canadian Journal of Economics*:329–48.

Helsley, Robert W., and William C. Strange. 1998. "Private Government." *Journal of Public Economics* 69 (2): 281–304.

Helsley, Robert W., and William C. Strange. 2014. "Coagglomeration, Clusters, and the Scale and Composition of Cities." *Journal of Political Economy* 122 (5): 1064–93.

Henderson, J. V. 1974. "The Sizes and Types of Cities." *American Economic Review* 64 (4): 640–56.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics* 11 (2): 1–39.

Isard, Walter, and Merton J. Peck. 1954. "Location Theory and International and Interregional Trade Theory." *The Quarterly Journal of Economics* 68 (1): 97–114.

Isard, Walter. 1954. "Location Theory and Trade Theory: Short-Run Analysis." *The Quarterly Journal of*

Economics 68 (2): 305–20.
Jacobs, Jane. 1961. *The Death and Life of Great American Cities* New York: Random House.
Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108(3): 577–98.
Jales, Hugo, Boqian Jiang, and Stuart S. Rosenthal. 2020. "Separating Selection from Spillover Effects: Using the Mode to Estimate the Return to City Size." https://ssrosent.expressions.syr.edu/research/.
Keller, Wolfgang. 2002. "Geographic Localization of International Technology Diffusion." *American Economic Review* 92 (1): 120–42.
Keller, Wolfgang. 2004. "International Technology Diffusion." *Journal of Economic Literature* 42 (3): 752–82.
Kerr, William R., and Scott Duke Kominers. 2015. "Agglomerative Forces and Cluster Shapes." *Review of Economics and Statistics* 97 (4): 877–99.
Konishi, Hideo, and Michael T. Sandfort. 2003. "Anchor Stores." *Journal of Urban Economics* 53 (3): 413–35.
Koster, Hans R. A., Ilias Pasidis, and Jos van Ommeren. 2019. "Shopping Externalities and Retail Concentration: Evidence from Dutch Shopping Streets." *Journal of Urban Economics* 114: Article Number 103194.
Li, Jing. 2014. "The Influence of State Policy and Proximity to Medical Services on Health Outcomes." *Journal of Urban Economics* 80: 97–109.
Liu, Crocker H., Stuart S. Rosenthal, and William C. Strange. 2018a. "The Vertical City: Rent Gradients, Spatial Structure, and Agglomeration Economies." *Journal of Urban Economics* 106: 101–22.
Liu, C., S. Rosenthal, and W.C. Strange. 2018b. "Building Specialization, Anchor Tenants, and Agglomeration Economies." https://ssrosent.expressions.syr.edu/research/.
Liu, C., S. Rosenthal, and W.C. Strange. 2020. "Vertical Density and Agglomeration Economies." Working paper.
Marshall, Alfred. 1890. *Principles of Economics* London: MacMillan.
Mas, Alexandre, and Enrico Moretti. 2009. "Peers at Work" *American Economic Review* 99(1): 112–45.
McMillen, Daniel P., and John F. McDonald. 1998. "Suburban Subcenters and Employment Density in Metropolitan Chicago." *Journal of Urban Economics* 43: 157–80.
McMillen, Daniel P., and Stefani C. Smith. 2003. "The Number of Subcenters in Large Urban Areas." *Journal of Urban Economics* 53 (3): 321–38.
Moretti, Enrico. 2004. "Workers' Education, Spillovers, and Productivity: Evidence from Plant-Level Production Functions." *American Economic Review* 94 (3): 656–90.
Moretti, Enrico. 2019. "The Effect of High-Tech Clusters on the Productivity of Top Inventors." NBER Working Paper 26270.
Ogawa, Hideaki, and Masahisa Fujita. 1980. "Equilibrium Land Use Patterns in a Nonmonocentric City." *Journal of Regional Science* 20 (4): 455–75.
Pashigian, B. Peter, and Eric D. Gould. 1998. "Internalizing Externalities: The Pricing of Space in Shopping Malls 1." *The Journal of Law and Economics* 41 (1): 115–42.
Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6): 1257–78.
Rosen, Sherwin. 1979. "Wages-based Indexes of Urban Quality of Life." In *Current Issues in Urban Economics,* edited by Peter M. Mieszkowski and Mahlon R. Straszheim, 74–104. Baltimore: John Hopkins University Press.
Rosenthal, Stuart S., and William C. Strange. 2001. "The Determinants of Agglomeration." *Journal of Urban Economics* 50(2): 191–229.
Rosenthal, Stuart S., and William C. Strange. 2003. "Geography, Industrial Organization, and Agglomeration." *Review of Economics and Statistics* 85(2): 377–93.
Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies." In *Handbook of Urban and Regional Economics,* Vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2119–72. Amsterdam: Elsevier.
Rosenthal, Stuart S., and William C. Strange. 2005. "The Geography of Entrepreneurship in the New York Metropolitan Area." *Federal Reserve Bank of New York Economic Policy Review* 11 (2): 29–53.
Rosenthal, Stuart S., and William C. Strange. 2008a. "The Attenuation of Human Capital Spillovers." *Journal of Urban Economics* 64 (2): 373–89.
Rosenthal, Stuart S. and William C. Strange. 2008b. "Agglomeration and Hours Worked." *Review of*

Economics and Statistics 90 (1): 105–18.

Rosenthal, Stuart S., and William C. Strange. 2012. "Female Entrepreneurship, Agglomeration, and a New Spatial Mismatch." *Review of Economics and Statistics* 94 (3): 764–88.

Tinbergen, Jan. 1962. *Shaping the World Economy: Suggestions for an International Economic Policy.* New York: Twentieth Century Fund.

Sandvik, Jason, Richard Saouma, and Nathan Seegert. 2019. "Workplace Knowledge Flows" https://faculty.utah.edu/u0908787-NATHAN SEEGERT/research/index.hml.

Vernon, Raymond. 1963. *Metropolis 1985: An Interpretation of the Findings of the New York Metropolitan Region Study.* New York: Doubleday.

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 50–76

Tech Clusters

William R. Kerr and Frederic Robert-Nicoud

While Silicon Valley houses less than 0.1 percent of the world's population, its shadow looms large. Many cities aspire to be a tech cluster: for example, an astounding 238 US cities jumped through hoops in 2017–18 to enter Amazon's infamous "bidding" process for where it would establish a second headquarters. Wikipedia lists more than 25 efforts to brand a US location as "Silicon Something," along with many foreign ones (at https://en.wikipedia.org/wiki/List_of_technology_centers#United_States). Our personal favorites are Silicon Peach (Atlanta) and Silicon Spuds (Idaho), whereas Silicon Prairie has at least four contenders. Other US examples include Silicon Anchor, Basin, Desert, Forest, Hill, Holler, Mountain, Shire, and Surf.

This paper examines the tech cluster phenomenon by considering three partially answered questions. We first ask how to define a tech cluster—that is, what properties are required to be a tech cluster? This delineation is harder than it appears at first glance and raises some key questions and issues. We start with the scale and density of local activity and then extend into the frontier nature of the work being undertaken and its ability to impact multiple sectors of the economy. We illustrate our definition through some common metrics like patents, venture

■ *William R. Kerr is the Dimitri V. D'Arbeloff—MBA Class of 1955 Professor of Business Administration, Harvard Business School, Boston, Massachusetts and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Frederic Robert-Nicoud is Professor of Economics, Geneva School of Economics and Management (GSEM), University of Geneva, Geneva, Switzerland, and Research Fellow, Centre for Economic Policy Research, London, United Kingdom. Kerr is the corresponding author at wkerr@hbs.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.50.

capital funding, and employment in sectors that are intensive in research and development or in digital-connected occupations. We also note some interesting clues from emerging metrics (for example, high-growth entrepreneurship, artificial intelligence researchers) and recent efforts to measure tech clusters globally.

We then ask how tech clusters function, with a focus on traits that extend beyond those associated with traditional industrial clusters. Not surprisingly, knowledge spillovers are a powerful force in tech clusters, and recent work explores how knowledge transmits across firms situated in a tech cluster and how density impacts the types of innovations created. Tech clusters facilitate powerful scaling for the best designs when they combine modular product structures with high-velocity labor markets. Universities, high-skilled immigration, and global production linkages also feature prominently in the functioning of leading US centers.

Finally, we turn to the roots of tech clusters and inquire into the mix of initial ingredients required for their formation. Leading tech clusters are far from permanent and have frequently emerged in new places following the advent of new general-purpose technologies. Today, the rapid growth of Toronto as an artificial intelligence cluster suggests that there may be limits to Silicon Valley's grip on this frontier. Yet despite the government having played an important role in this history of many tech clusters, top-down attempts to re-create Silicon Valley have mostly failed (Lerner 2009). Our historical examples suggest that local officials instead may wish to facilitate the scaling of nascent industries that have taken root, even if due to random chance, rather than attempt to engineer a cluster from scratch.

We conclude with some thoughts on future research opportunities, including the question of whether tech clusters are at their high-water mark or are likely to strengthen further. The implications of the ongoing COVID-19 crisis for tech clusters could be profound. Our discussion focuses primarily on the US economy, but much of what we describe applies to other countries as well. We ground our discussion firmly within the economics and management disciplines, occasionally reaching out in incomplete ways to other social sciences as we go.

Defining Tech Clusters

While it is easy to point to high profile examples of tech clusters, such as Seattle or Austin, developing even a semi-formal definition is tricky. "Clusters" traditionally indicate an important overall scale of local activity, complemented by spatial density and linkages amongst local firms (for example, Marshall 1890; Porter 1998). As discussed further below, the specific linked activities for tech clusters might include engineer mobility across employers, flows of technical knowledge, and reliance on shared local inputs like a research university. In addition to these traditional dimensions, we define "tech" clusters to be locations where new products (be they goods or services) and production processes are created that affect multiple parts of the economy. That is, a tech cluster must have a frontier edge, and it must extend beyond refinements to a single industry.

These criteria suggest that tech clusters are not a new phenomenon nor a permanent fixture. Indeed, US economic history shows a continual movement of leading tech centers: for example, Lowell, Massachusetts, for textile mills reliant on water power in the 1800s; Cleveland, Ohio, for electricity and then steel in the early 1900s; and Detroit, Michigan, for automobiles in the early–mid 1900s (Lee and Nicholas 2013; Lamoreaux, Levenstein, and Sokoloff 2004). Our definition puts early nineteenth century technology advances for engines in Detroit on par with the cluster of artificial intelligence firms in the Toronto area today, which seems conceptually useful.

A historical perspective also suggests that tech clusters may cease to be. For example, Detroit was the Silicon Valley of the first half of the twentieth century. At some point, the auto industry matured and Detroit with it, and we would have taken away Detroit's tech cluster badge. Should Detroit's mojo return with electric or autonomous vehicles—perhaps in 2030—we would declare Detroit a tech cluster again. Over its relatively short history, Silicon Valley has also experienced doldrums after specific technology waves crested and before the next major path emerged.

Our definition also suggests drawing a line between specific industries which make heavy use of technology (which includes traditional industrial districts), and a true tech cluster with a broader impact across the economy. For example, should Wall Street and the surrounding area of lower Manhattan be considered a tech cluster? After all, Goldman Sachs in 2020 employs more engineers than the total combined workforces of LinkedIn and Twitter. The iconic bank recently has even been shedding traditional practices like dress codes to attract technical workers. Frontier quantitative hedge funds are at the bleeding edge of artificial intelligence, and fintech advances may reshape commerce. Maybe the Wall Street of the 1980s was not a tech cluster, but the Wall Street of 2030 might be. Using the framework of Duranton and Puga (2005), perhaps Wall Street is evolving from being a cluster specialized in a sector—financial services—into a cluster specializing in a function—(fin)tech.

These definitional challenges reflect how advanced technology and its leading firms are entering many parts of the economy in a variety of ways. Technology is becoming less of a segmented industry—for example, less focused on manufacturers of personal computers or shrink-wrapped software—and more of a ubiquitous and general purpose one. There also exists a blurring of industry boundaries, especially as incumbent firms seek to move out of stagnating industries and towards new profitable opportunities. As robotics and cognitive automation advance, this ambiguity will grow. Technology is becoming so pervasive that one can be tempted to resort to phrasings like "talent clusters" to focus on frontier activity by sector in human-capital focused industries (for example, Kerr 2019).

Data to Measure US Tech Clusters

The empirical study of tech clusters requires making choices about what to measure and the appropriate scale of activity. Most analyses use patents, high-growth entrepreneurship supported by venture capital firms, and/or employment

in industries or occupations that are intensive in research and development. In choosing a geographic unit, most empirical analyses of the US economy examine the full distribution of states or cities, which is helpful for getting a workable sample size (for example, Acs, Anselin, and Varga 2002; Delgado, Porter, and Stern 2010; Glaeser, Kerr, and Kerr 2015). An alternative method is to conduct case studies or sub-city empirical analyses of a recognized tech cluster like Silicon Valley (for example, Saxenian 1994; Kenny 2000; Bresnahan and Gambardella 2001). These choices should follow the type of economic linkage under study: for example, focusing on very short-distance knowledge spillovers in the area around Kendall Square near MIT versus the labor mobility of engineers across the entire Boston metropolitan area.

Patents and venture capital data are popular with researchers due to the existence of detailed micro-data regarding individual inventions and funding transactions. Thus, in addition to measuring spatial concentration, researchers can use the same data to learn how the clusters operate—for example, by following the careers of inventors or entrepreneurs over time, modeling local networks and spillovers, and so on. These data also offer a foothold for assessing whether the innovative work of the city touches multiple aspects of the economy. The central liability focusing on patents and venture capital data is that many forms of innovative activity are not captured; moreover, the intellectual property and financing environment changes over time (for example, as a result of greater recognition of software or business method patents). Researchers must carefully consider comparability across industries (and therefore across cities, too) and longitudinally (see literature in Feldman and Kogler 2010; Carlino and Kerr 2015).

With some exceptions, such as Carrincazeaux, Lung, and Rallet (2001) and Carlino, Carr, and Smith (2012), location-specific data on research and development are difficult to acquire. Industry- and occupation-level employment data offer another tactic. As an example, we use below micro-data from the 2014–2018 American Community Survey that records for individuals their metropolitan area, industry of employment, salary, education level, and so forth. We map research and development intensity by industry (as documented by the National Science Foundation 2017) to measure how much of a city's employment base is in R&D-intensive fields. This approach avoids some of the liabilities noted for patenting and venture data but also sacrifices many of the advantages that micro-data provided.

Table 1 documents several measures for cities using data from around 2015–2018 (the notes to the table provide details on sources and preparation). We list the top 15 metropolitan statistical areas in terms of venture capital investment in descending rank and then provide two aggregate categories for the other 266 metropolitan statistical areas and for rural areas. In this table and the figures to follow, we use consolidated metropolitan statistical areas, such that the San Francisco/San Jose/Oakland area is simply referred to as San Francisco.

This table speaks best to the scale of tech activity across cities, and through a comparison to the population share in the final column, the implied density of tech efforts. The top 15 metropolitan statistical areas as ranked by venture capital

Table 1
Spatial Concentration of US Tech Activity

Consolidated metro area	*Venture capital investment*	*Granted patents*	*Employment in top 10 R&D industries, high-skilled*	*Employment in top 20 R&D industries, all workers*	*Employment in computer- and digital-connected occupations, high-skilled*	*Employment in STEM-connected occupations, all workers*	*Population*
San Francisco	48.1%	18.4%	11.7%	4.9%	8.6%	5.5%	2.5%
New York	15.3%	6.0%	6.3%	5.1%	8.0%	6.0%	6.4%
Boston	10.5%	4.5%	5.5%	2.4%	3.4%	2.7%	1.6%
Los Angeles	6.5%	5.3%	5.6%	5.7%	3.9%	3.9%	5.8%
Seattle	2.1%	4.0%	4.2%	2.4%	3.5%	2.5%	1.2%
San Diego	1.9%	3.6%	3.2%	1.6%	1.5%	1.5%	1.0%
Chicago	1.7%	2.5%	3.2%	3.2%	3.9%	3.2%	2.9%
Washington DC	1.5%	1.7%	4.4%	1.8%	6.6%	4.6%	1.8%
Miami	1.5%	0.7%	0.9%	1.1%	1.0%	1.2%	1.4%
Denver	1.1%	1.5%	1.5%	0.9%	1.7%	1.5%	1.0%
Austin	1.0%	2.1%	1.8%	1.0%	1.5%	1.2%	0.6%
Philadelphia	0.8%	1.8%	3.3%	2.1%	2.4%	2.2%	2.0%
Atlanta	0.7%	1.5%	1.4%	1.6%	2.8%	2.3%	1.7%
Minneapolis-St. Paul	0.7%	2.0%	1.3%	1.7%	2.0%	1.9%	1.0%
Raleigh-Durham	0.5%	1.4%	1.7%	0.8%	1.2%	1.0%	0.5%
Share in top 15 VC MSAs	93.8%	57.0%	55.9%	36.0%	52.1%	41.2%	31.3%
Share in other MSAs	5.9%	37.3%	38.3%	49.3%	41.8%	47.9%	48.0%
Share in non-metro areas	0.3%	5.7%	5.9%	14.8%	6.1%	10.9%	20.7%
Correlation to VC share		0.98	0.91	0.63	0.73	0.66	0.31
Correlation to patent share	0.98		0.93	0.67	0.71	0.65	0.32

Note: Table lists the top 15 (consolidated) MSAs in terms of venture capital investment in descending rank. Venture capital investments are for 2015–2018 based upon location of new investments in ventures and are taken from Thomson One. Patents are for 2015–2018 based upon the most frequent location of inventors, and application date of utility patents are taken from patents granted by the USPTO through the end of 2019. Employment columns are for 2014–2018 using the combined American Community Survey (ACS) 1% files. ACS sample includes those aged 18–65 who are working and with positive wage earnings, not in group quarters, with usual hours worked greater than 30 per week, and with usual weeks worked per year greater than 40. High-skilled workers are those with college degrees or higher in education and earn $50,000 or more. The ten industries with the highest R&D per worker as listed by NSF (2017) are Software publishers; Pharmaceuticals and medicines; Other computer and electronic products; Data processing, hosting, and related services; Communications equipment; Semiconductor and other electronic components; Navigational, measuring, electromedical, and control instruments; Pesticide, fertilizer, and other agricultural chemicals; Aerospace products and parts; and Scientific research and development services. These industries in some cases map onto more than one NAICS industry in the ACS for employment data. Population data are from 2015–2018 based upon counties that comprise MSAs and are taken from the Census Bureau. There are 281 MSAs identified in the venture capital, patent, and population data and 261 identified in the ACS data. Population distributions in the ACS are very similar, with the one noticeable difference of LA being a 4.2% share.

investment hold 94 percent of venture capital activity in the first column and 57 percent of patenting in the second column, compared to just 31 percent of the population. If we instead rank on patents, Detroit, Portland, Dallas-Ft. Worth, and Houston feature in the 15 largest centers, with Washington, Miami, Atlanta,

and Raleigh-Durham dropping out. Either way, patenting and especially venture capital investment are underrepresented outside of leading tech centers. Looking across the metro areas listed in Table 1, shares for venture capital and patents have a 0.98 correlation, while shares for venture capital and population have a 0.31 correlation.

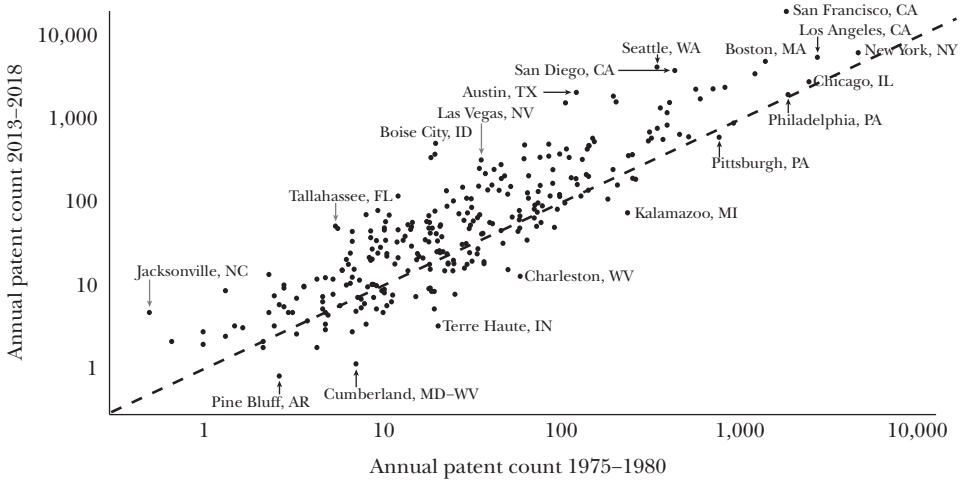
The third and fourth columns of Table 1 provide two measures of local employment in leading industries for R&D investment as measured by National Science Foundation (2017). We first show a restrictive definition, where we identify college-educated workers earning more than $50,000 (short-hand labelled as "high-skilled") and working in a top 10 R&D-intensive sector—11.7 percent of such individuals work in the San Francisco area, compared to 5.9 percent of them being outside metropolitan areas. The second measure broadens to any full-time employee (no education or salary restriction) among the 20 most R&D-intensive sectors. This makes a noticeable difference, with San Francisco's share now 4.9 percent and much smaller than the 14.8 percent in non-metro locations. The fifth column looks first at high-skilled workers in occupations in computer- and digital-connected work, and the sixth column expands to all full-time workers in a broader class of STEM-connected occupations.

This table shows the potential and challenges of defining tech clusters using the scale and density of local tech activity. Six cities appear to qualify under any aggregation scheme: San Francisco, Boston, Seattle, San Diego, Denver, and Austin all rank among top 15 locations for venture capital and for patents (scale) and hold shares for venture capital, patents, employment in R&D-intensive sectors, and employment in digital-connected occupations that exceed their population shares (density). They also pass our highly rigorous "sniff test"—that is, they just make sense. Washington, Minneapolis-St. Paul, and Raleigh-Durham would join the list if relaxing the expectation that the share of venture investment exceed population share (which is hard due to the very high concentration in San Francisco).

New York and Los Angeles are more ambiguous: they hold large venture capital markets (and venture investors frequently declare them leading tech clusters), but their patents and employment shares in key industries and fields are somewhat less than their population shares. Were we to disaggregate these huge metro areas, we would likely identify a sub-region that would independently fit on this short list by still holding sufficient scale and yet having achieved a more recognizable density. Said differently, there is surely a part of New York and Los Angeles that would be stand-alone equal to or greater than Austin (for example, Egan et al. 2017). Chicago's activity is mostly equal to its population share or less.

At the other end of the city size distribution, it is hard to be a robust yet small tech cluster on both venture investment and patent metrics due to the concentration of innovation. If one only requires that a tech cluster achieve a venture capital and patent share that is 1.5 times the local population share, the one new city would be Provo, Utah, with Denver dropping out. In summary, San Francisco and Boston are extreme cases, and we are probably looking at 5–10 additional leading centers across the country depending upon definition of scale and density.

Figure 1
Growth in Annual Patenting by Metropolitan Statistical Area



Note: Figure presents for metropolitan areas the average annual patent count for 1975–1980 and 2013–2018. Patents are grouped by application year and all patents granted by the USPTO through end of 2019 are used. Axes are in log format and a 45-degree line is included. Some cities are labelled for illustrative purposes only.

At the start of this section, we conceptualized tech clusters as being positioned in frontier sectors and having a broad-based impact. Patents provide a preliminary example of these traits. We first consider new technology areas by isolating patent technology classes that the US Patent and Trademark Office introduced in 1995 and afterwards. On average, cities have 7.8 percent of their patents during 2015–2018 in the newest classes, while the average for San Francisco, Boston, Seattle, San Diego, Denver, and Austin is 27.8 percent. When looking at patent classes introduced after 1980, these shares are 29.8 percent and 60.2 percent, respectively. Patents in these six cities also display higher forward and backward citations, with a greater measure of generality to the work (Hall, Jaffe, and Trajtenberg 2001). We return below to recent research describing differences in the type of innovation across clusters.

How is this picture changing over time? For the most part, the rich are getting richer. Figure 1 shows city patenting (presented in annual terms) from 1975 to 1980 and from 2013 to 2018. The axes are in log format and a 45-degree line is included. There has been an overall increase in patent grants since the late 1970s, visible in the figure with more cities being above the 45-degree line than below. Cities that are farthest above the 45-degree line have the biggest percentage gains, and big patenting centers in the late 1970s show the most consistent increases. Consequentially, an Ellison and Glaeser (1997) index of patenting concentration relative to

Table 2
Global Tech Clusters as Measured by Total Size

Venture Capital Investment (Thomson One, 2009–2018)	*Unicorn Startup Companies (CB Insights, 2009–2018)*	*Patent Cooperation Treaty Filings (WIPO, 2010–2015)*
San Francisco	San Francisco	Tokyo-Yokohama
Beijing	Beijing	Shenzhen-Hong Kong
Shanghai	New York	San Francisco
New York	Los Angeles	Seoul
Boston	Shanghai	Osaka-Kobe-Kyoto
Los Angeles	Boston	San Diego
London	London	Beijing
Shenzhen	Seattle	Boston
San Diego	Hangzhou	Nagoya
Seattle	Chicago	Paris

Note: Table lists the 10 largest global tech clusters in terms of various metrics in descending rank. Venture capital investments are for 2009–2018 based upon the location of new investments in ventures and are taken from Thomson One. Unicorn startup companies are counts of new ventures exceeding a billion dollars in valuation during 2009–2018 and are taken from CB Insights. Patent Cooperation Treaty filings are for 2010–2015 and are taken from the World Intellectual Property Organization. Geographic boundaries of clusters are defined by each data source and differ to some extent across columns.

population distribution grows over ten-fold from an index value of 0.002 in the late 1970s to 0.028 in 2018.

Researchers have recently developed new empirical methods to measure tech clusters, as well. One approach focuses on measuring high-growth entrepreneurship independent of venture capital data. Guzman and Stern (2019) use state-level business registration data and develop techniques to identify whether new firms are targeting rapid growth, such as how the venture is named (for example, Infinity Global Technologies versus Fred's Bicycle Repair) and its legal form of incorporation. The most intense areas for entrepreneurial potential are places like Silicon Valley, Boston, and Austin, where they also measure booms in local high-growth activity through 2019. In another approach, using LinkedIn data on employment, Gagne (2019) estimates that more than one-third of artificial intelligence researchers are located in the San Francisco Bay Area—a fact due in part to the presence of tech giants like Microsoft, IBM, and Google in that area.

Global Tech Clusters

An emerging frontier is to map out global tech clusters. This combination of data across borders gets complicated fast, and Table 2 shows that metrics do not always point in the same way. For venture capital investment, the last decade shows the remarkable rise of Chinese tech clusters. The top ten global cities include Beijing, Shanghai, and Shenzhen, plus London, in addition to six cities from the United States. Looking instead at the post 2009 formation of unicorn start-ups (valued at $1 billion or more), the four non-US cities are similarly Beijing, Shanghai, and Hangzhou, plus London (Kerr 2018).

While measures of tech clusters using venture capital and patents provide mostly similar pictures across US cities, globally this is not the case. In a World Intellectual Property Organization report (Bergquist, Fink, and Raffo 2017) that aggregates over many patent offices, Tokyo-Yokohama holds twice the patent count to second place, Shenzhen-Hong Kong; the San Francisco Bay Area is third and Seoul is fourth. Moreover, the top ten cities span three in Japan, three in America, two in China, and one each in Korea and France. For more specific frontiers like research in artificial intelligence, the leading roles of America and China are clear, but relative shares depend substantially on the yardstick employed and data source.

Building a stronger foundation for these comparisons is an important ongoing task. So far, we are only tackling the scale of local tech activity but not the extra nuances about density, frontier status, and so forth. International settings also raise the interesting question of whether measures of a tech center should be context specific. Many speak of Bangalore as a "tech cluster," but while that area is technologically advanced when compared to other locations in India, much of its activity is substantially lower tech and labor intensive relative to tech clusters in advanced economies.

Is a Tech Cluster Different from Other Clusters?

Industry clusters arise due to the production advantages of local specialization combined with subsequent trade across locations. Marshall (1890) famously described three forces of what we now call agglomeration economies: knowledge spillovers, labor market pooling, and customer-supplier interactions. Economic research over the last two decades has shown all three forces, along with natural advantages of areas for certain industries (like harbors or coal mines), are important for explaining industrial clusters, with the most recent research quantifying the heterogeneity across industries and co-agglomeration dynamics over time (for example, Ellison, Glaeser, and Kerr 2010; Faggio, Silva, and Strange 2017). While most studies of the Marshallian forces have focused on industrial settings, they also apply to tech clusters and often in distinctive ways.[1]

Knowledge Spillovers and Forms of Innovation

Our definition of tech clusters emphasizes settings with a frontier edge, and many companies seek insights on emerging possibilities, either through first access to codified knowledge or to tacit knowledge that cannot be written down easily. Marshall (1890) famously described knowledge diffusion inside an industrial cluster in poetic terms: "The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously."

[1] Duranton and Puga (2004) recast Marshall's forces to emphasize higher-order functions like sharing and matching that occur within clusters. See Markusen (1996) and Porter (1998) for complementary approaches.

Researchers have since catalogued these knowledge transfers in many settings, such as Switzerland's watchmaking industry, and they appear particularly important for tech clusters (Audretsch and Feldman 1996). Olson and Olson (2003) document very tight bands for collaborative interactions. In an ethnographic study of Silicon Valley, Saxenian (1994, p. 33) describes many formal and informal channels facilitating knowledge transfer, including a depiction of Wagon Wheel, a Mountain View bar that novelist Tom Wolfe dubbed the "fountainhead of the semiconductor industry":

> [M]embers of an 'esoteric fraternity'—the young men and women of the semiconductor industry—would head after work to have a drink and gossip and brag and trade war stories about phase jitters, phantom circuits, bubble memories, pulse trains, bounceless contracts, burst modes, leapfrog tests, p-n junctions, sleeping sickness modes, slow-death episodes, RAMs, NAKs, MOSes, PCMs, PROMs, PROM blowers, PROM blasters, and teramagnitudes, meaning multiples of a million millions.

More recently, then-CEO Jeff Immelt (as reported in Singer 2016) described why General Electric was moving its headquarters from Fairfield, Connecticut, to Boston, Massachusetts: "To look out the window [in Connecticut] and see deer running across, I don't care about that. I want some 29-year-old [graduate of] MIT to punch me right in the nose and say all of GE's technologies are wrong and you're about to lose. That's the challenge." Kerr (2018) discusses the subsequent ups and downs of General Electric's move.

More formally, economists since Jaffe, Trajtenberg, and Henderson (1993) have most frequently used patent citations to quantify the higher rate of knowledge flow within cities versus across them (for example, see Murata et al. 2014, and the references cited therein). The use of patent citations is only an imperfect proxy for knowledge flows (for example, Jaffe, Trajtenberg, and Fogarty 2000), and many of the information flows captured by patent citation data are surely due in practice to inventor networks, licensing agreements, and so forth (for example, Almeida and Kogut 1999; Breschi and Lissoni 2009). These citation metrics thus aggregate unpriced knowledge spillovers that are "in the air," alongside regular forms of economic activity. Citation patterns have been confirmed with co-authorship networks among inventors, and Fleming and Marx (2006) identify that leading tech clusters became more connected during the 1990s.

Another use of patent data is to open the black box of how tech clusters operate. Kerr and Kominers (2015) model localized spillovers within tech clusters. Firms interact with their closest neighbors, but the costs of interaction prevent direct spillover benefits from more distant members of the cluster. For example, a firm in Oakland may have useful information for a startup in East Palo Alto, but the search and acquisition costs for that information prevent it from diffusing directly, requiring, instead, indirect transfer via other firms. These conditions lead to overlapping zones of interaction, such that nearby interactions are direct, while those

farther away happen through the underlying network of the cluster. Arzaghi and Henderson (2008) document a similar phenomenon in a study of advertising agencies in Manhattan.

In their empirical work using patent citations, Kerr and Kominers (2015) show that firms are more likely to cite directly the work of their closest neighbors but to cite indirectly those farther away in the cluster. Consequently, econometricians can compare the shapes and sizes of clusters to learn about the technologies that sit behind them. Technologies with tight spillover lengths produce smaller and denser clusters. In this study as well as other research using broader sources of variation (for example, Rosenthal and Strange 2001, 2003), knowledge spillovers are the most localized of agglomeration forces.[2]

Another promising line of work quantifies how the level and type of inventions varies within a broader metro area. For example, Carlino, Chatterjee, and Hunt (2007) and Berkes and Gaetani (2019) find that patenting per capita across US cities mostly rises with higher population density, with a 10 percent increase in density correlating with a 2 percent increase in intensity. At a more fine-grained level, however, patenting per capita peaks in areas with high but not too high density—for example, being higher in Silicon Valley or the Route 128 area surrounding Boston compared to downtown San Francisco and Boston, respectively.

Berkes and Gaetani (2019) further show that the very densest districts instead foster atypical combinations of technologies that combine core elements seen in prior work with distinctly novel elements (Uzzi et al. 2013). These innovation advantages for developing the most novel forms of new work are often credited to a diverse range of local inputs (for example, Jacobs 1970; Glaeser et al. 1992; Henderson Kuncoro, and Turner 1995; Lin 2011). In contrast, "company towns" where a single large firm dominates the local tech activity, like Eastman Kodak in Rochester, New York during the middle of the twentieth century, are more likely to have internally focused innovation (Agrawal, Cockburn, and Rosell 2010).

Continued investigation into how the technologies developed in frontier clusters differ from other settings is important. It would be interesting as well to identify cases and situations in which tech clusters can become too isolated from a potential customer group to understand latent needs. Michael Bloomberg is a very rich tech entrepreneur because he knew what kinds of desktop terminals his former colleagues on Wall Street were missing, which someone in California may have had a hard time figuring out.

[2]Even controlling for distance, political boundaries still matter for knowledge flows (Singh and Marx 2013). Similarly, local economic conditions (low commuting costs, skilled labor abundance) and technology features (localized knowledge spillovers, high startup costs) shape the decentralized emergence of science parks (Liang et al. 2019). By contrast, some studies do not find co-location to be essential (for example, Waldinger 2012).

Specialized Labor and High Velocity Labor Markets

A distinctive feature of tech clusters is the specialized skill sets of many local workers, which then become a powerful magnet to the area. As noted earlier, leading tech clusters hold a large share of the nation's college-educated workforce engaged in computer and digitally connected fields, and the concentrations become even more skewed when looking at extreme skills like specialization in artificial intelligence (for example, Gagne 2019). Clusters provide several advantages for workers with specialized skills: insurance against the shocks befalling any one employer, deeper labor markets for better matching of particular skill sets with the best jobs, and often superior environments for investments in training by talented individuals without fear of later employer hold-up (for an entry point to this literature, see Overman and Puga 2010, and the citations therein). Studies examining labor pooling in the tech arena often emphasize its role for employee-firm matching and input sharing (for example, Helsley and Strange 1990, 2002).

Beyond these bread-and-butter features, the literature on tech clusters most often emphasizes the high velocity turnover of its labor markets. Saxenian (1994, p. 35) provides an early depiction of this rapid mobility, quoting an engineer on the ease of transitioning employers in Silicon Valley: "Out here, it wasn't that big of a catastrophe to quit your job on Friday and have another job on Monday and this was true for company executives. You didn't necessarily even have to tell your wife. You just drove off in a different direction on Monday morning. You didn't have to sell your house, and your kids didn't have to change schools." Another local executive notes: "People change jobs out here without changing car pools."

High profile executive moves are common within tech clusters, such as Sheryl Sandberg's move from Google to become Chief Operating Officer of Facebook in 2008 and Marissa Mayer's similar departure to become CEO of Yahoo! in 2012. These moves often spark legal challenges. In 2017, Alphabet's Waymo sued Uber, alleging that one of Waymo's former engineers, Anthony Levandowski, took confidential files with trade secrets related to self-driving cars with him when leaving to form his own self-driving startup, Otto, that Uber later acquired. The suit was settled in 2018 with Uber paying 0.34 percent of its equity (then valued at $245 million) to Waymo (as reported in Marshall 2018).

While the velocity of these labor transitions has been frequently discussed, it has been less studied empirically compared to the localization of knowledge flows. Fallick, Fleischman, and Rebitzer (2006) is an important exception that further links the flexible labor markets of tech clusters to an industrial organization that emphasizes modular production.[3] They model how modularity allows for winner-take-all competition, with labor rapidly reallocating to the firm with the best design

[3]Modularity is the method of making complex products or creating processes from smaller subsystems developed by a network of independent firms. Although different suppliers are responsible for separate modules, they follow "design rules" that ensure the modules work together (Baldwin and Clark 1997). This approach decentralizes innovation and may accelerate technical progress, since independent firms can focus innovation to their specific components compared with the divided attention of vertically integrated firms. Saxenian (1991), Sturgeon (2002), and Berger (2005) provide case examples.

in order to scale it up for production. This benefit helps the cluster to overcome potential underinvestment in worker training due to rapid turnover in high-velocity labor markets. Related, Gerlach, Rønde, and Stahl (2009) connect labor pooling to greater risk taking with research and development activities inside tech clusters. Fairlee and Chatterji (2013) document how rapid scaling of winning firms can ironically reduce start-up rates inside tech clusters during exceptional growth periods like the late 1990s.

This rapid labor mobility hints at the dual-edge nature of tech clusters; while they provide strong advantages, they impose real costs on firms, too. Despite the relative abundance of sought after skills within tech clusters, these labor markets were exceptionally tight in the late 2010s and exhibited very low unemployment rates. Thus, many businesses located in these talent clusters struggled to get the workers they wanted especially if they lacked a brand name like Apple or Netflix that attracts employees.

Firms also need to be aware that company doors operate in both directions. While bosses get excited about the top-notch employees and knowledge stocks at neighboring companies that they might be able to lure away, they also become more likely to have their own employees depart to rival organizations. Combes and Duranton (2006) model this tension, showing that single-minded pursuit of a position in the cluster is not always the best strategy. Building on Rotemberg and Saloner (2000), Matouschek and Robert-Nicoud (2005) and Almazan, De Motta, and Titman (2007) highlight the role of firm-sponsored investments and firm-specific skills in investigating why employers should think twice before jumping into the hot spot of their sector. Alcácer and Chung (2007, 2014) and Groysberg (2010) consider these themes in the management literature.

These tensions stress how clusters are an outcome of an equilibrium process. Thus, places with great spillover benefits usually bring very high prices for real estate and talent. This market pricing is true across cities and across small zones inside prominent clusters. Not only is Boston more expensive as a whole than Providence, the real estate around Kendall Square and MIT is the priciest. Indeed, abstracting from moving costs, escalating real estate prices can enhance the fidelity of the cluster, as only those who most benefit from the location are willing to pay astronomical rates (for example, Malmberg and Power 2006; Bathelt and Li 2014). Few studies have explicitly modelled these tradeoffs and tensions, and yet they are critical for our understanding.

These labor tensions extend into employment law. Non-compete clauses in employment contracts limit the ability of a person to leave their employer and immediately compete in the same segment. Gilson (1999) proposed that Silicon Valley's dynamism should be attributed to the inability of local firms to enforce non-compete clauses. While non-compete clauses may encourage employers to invest more in training workers, as they are less likely to be poached by rivals, the labor rigidities can also stifle the flow of ideas and the optimal matching of workers and firms. Subsequent empirical analyses by Marx, Strumsky, and Fleming (2009), Marx, Singh, and Fleming (2015), and Hausman (2019) have

shown such rigidities to be particularly troublesome for inventors and technical diffusion.[4]

Immigration, Diversity, and Tech Talent

Immigration and talent diversity, two factors not discussed by Marshall (1890), are also critical for the understanding of US tech clusters. Classic early accounts of tech clusters by Saxenian (1994), Saxenian, Motoyama, and Quan (2002), and Florida (2005) emphasize how openness and tolerance in the community undergird the innovative productivity of the cluster. These authors, along with Falck, Fritsch, and Heblich (2011), further consider how urban amenities and high quality of life are necessary to attract the highly skilled people central for tech clusters.

US tech clusters are high-skilled immigration hubs, in most cases building on strong past waves of immigration to large coastal cities. More than 60 percent of Silicon Valley's entrepreneurs are immigrants to America (Kerr and Kerr 2020), and the chief executive officers of Alphabet, Microsoft, SpaceX/Tesla, and Uber are all foreign-born. Much of the large innovative workforce of tech clusters comes from abroad. Immigrants accounted for an astounding two-thirds of the college-educated workforce in San Jose, California, in the American Community Survey for information and communications technologies. While San Jose is an outlier with its location in Silicon Valley, immigrants as a share of the college-educated workforce in these fields still exceed 40 percent in many tech clusters.

Kerr (2019) describes factors behind this reliance: talent for science, technology, engineering, and mathematics is quite transportable across countries, and the ranks of foreign talent looking for education and subsequent work opportunities in America in tech fields has been growing, especially from China and India. Part of America's immigration system is employer-driven (as a prominent example, the H-1B temporary visa program for those in "specialty occupations"), which also offers technology firms a substantial lever for using foreign talent. Not surprisingly, a literature has quantified how growth in US immigration can benefit tech clusters and their major employer firms (for example, see Kerr and Lincoln 2010; Peri, Shih, and Sparber 2015). Nathan (2014, 2015) provides similar evidence with a European focus.

A distinguishing feature of tech clusters is their cultural celebration of innovation that has the potential to change the world. But other common cultural forces in tech clusters can be counterproductive. Contrary to the growing evidence of a diversity premium for generating ideas, tech clusters have been frequently plagued by a "bro" culture that disadvantages women and minorities. Despite high-profile tech leaders like Mayer and Sandberg, women are underrepresented and sometimes

[4]Firms can also seek extra-legal maneuvers. In the late 2000s, major tech employers entered into anti-poaching agreements with each other, later paying large fines to settle the cases (as reported in Roberts 2015; Mehrotra 2016).

dramatically so (for example, only 2–3 percent of venture funding goes to women entrepreneurs). African-American participation is also terribly low, with recent gains in professional occupations like management consulting and investment banking not occurring in tech work (Gompers and Wang 2017). A separate concern is that tech companies may still operate with the "move fast and break things" spirit, but broader public concerns regarding privacy, data security breaches, and propagation of "fake news" via social media loom large.

Customer-Supplier Interactions, Firm Organization, and Global Networks

Returning to the last of Marshall's forces, the benefits that firms in tech industries gain from co-locating depend upon local production techniques and, perhaps less obviously, on global integration and production chains. Taking the local perspective first, many case examples point to the critical nature of local supply (Saxenian 1991). An early Apple executive described the desire for regional proximity: "Our purchasing strategy is that our vendor base is close to where we're doing business. . . We like them to be next door. If they can't, they need to be able to project an image like they are next door." Even where manufacturing was to be ultimately off-shored, contract manufacturer Flextronics emphasized local integration: "In the early stage of any project, we live with our customers and they live with us. Excellent communication is needed between design engineers, marketing people, and the production people, which is Flextronics."

Agrawal and Cockburn (2003) and Feldman (2003) developed concepts of "anchor firms" for clusters, which all those cities hoped to achieve by luring Amazon's HQ2 to their area, and Glaeser and Kerr (2009) considered optimal industrial composition. Markusen (1996) and Agrawal et al. (2014) emphasize the importance of firm size diversity. Large local firms anchor the cluster and produce ideas that do not fit well internally and thus get spun-out. Many small firms are also vital to lower entry barriers and to stimulate specialized support services. This local diversity was present in Detroit in the early 1900s and Silicon Valley in the 1960s (for example, Klepper 2010), and Agrawal et al. (2014) find evidence for their model when looking at the innovative output of US cities during the 1975–2000 period.

Hellmann and Perotti (2011) alternatively conceptualize how tech clusters facilitate the generation, circulation, and completion of new ideas. They model an important tradeoff of seeking to circulate and complete novel ideas within firms (where they are more protected) versus in local clusters (where they are more likely to find best matches). Their model predicts diverse organizational forms—internal ventures, spin-offs, and start-ups—coexisting and mutually reinforcing each other. An empirical analysis of these features, along with the acquisition of ideas into firms, seems very promising for future research.

While the economics literature mostly studies the local properties of tech clusters, they must also be embedded in the larger value chain of an industry (Coe and Bunnell 2003; Humphrey and Schmidt 2002). While Apple and Google race to design the next features of the smart phone, for example, the phones themselves are

produced in much lower cost locations and sold in retail shops globally. The geography literature discusses how tech clusters achieve their scale by integrating the local "buzz" into regional, national, or global production networks (for example, Storper and Venables 2004; Bathelt, Malmberg, and Maskell 2004; Bathelt and Li 2014). In addition to allowing rapid local scaling, modular production design makes it easier for supply chains to extend across multiple locations and over borders.

The linkages between global tech centers are also important and growing. In addition to constituting a large share of the local innovative workforce, high-skilled migrants facilitate many exchanges between tech centers (Saxenian, Motoyama, and Quan 2002; Saxenian 2007), and a substantial share of patent inventor teams are now cross-border (Miguelez 2014; Branstetter, Li, and Veloso 2015; Kerr and Kerr 2018). Venture capital firms are especially well connected internationally (Balachandran and Hernandez 2019), and leading corporations maintain a string of labs and move workers between facilities (Choudhury 2016, 2017). Nanda and Khanna (2010) also emphasize the degree to which time abroad can aid entrepreneurs when they return to less well-connected parts of their home country.

Preconditions and Dynamics of Tech Clusters

An emerging frontier of research focuses on whether tech clusters can be created, and the necessary preconditions in doing so, with a persistent meta-finding that it is very difficult to predict where leading clusters will take root. Krugman (1991) emphasizes the role of historical accidents in explaining *where* clusters form and how local efforts to "become the next Silicon Valley" have a poor track record (see discussions and references in Lerner 2009; Duranton 2011; Chatterji, Glaeser, and Kerr 2014). Though history provides multiple examples of the development of a new tech cluster, predicting or purposefully creating the location of the next cluster might be impossible.

For example, in a portrait of the origins of Silicon Valley, Lee and Nicholas (2013) note that San Mateo County was a technological backwater for several decades from the 1890s. It was not until the 1930s that the area began to be noticed for its work on transistors, vacuum tubes, and microwaves, which helped draw in larger firms and enabled startups. The government's huge demand for electronics in World War II brought critical mass to the region, as the local population of tech engineers surged ten-fold in a few years. When Silicon Valley went through its inflection point, many other cities would have looked much better prepared in terms of industry composition and talent base to be the next leading center. Indeed, accounts of the formation of Silicon Valley like Saxenian (1994) emphasize how the region's "blank slate" allowed for new forms of work to emerge, versus some preexisting factor that destined the region for success. Being a "blank slate" may have worked for Silicon Valley, but it is not a strategy that consistently guarantees success!

In most accounts of the origin of tech clusters, such as Klepper's (2010, 2016) comparisons of Detroit and Silicon Valley, emphasis is given to the initial placement

of a few important firms and the spinoff companies they subsequently generate. This outsized influence for anchor firms generates ample room for random influences on the early location decisions vital to a future cluster. For example, William Shockley, who shared a Nobel Prize in Physics for his work on semiconductors and transistors, moved to the San Francisco area to be near his ailing mother. Later, the spinoffs from his firm Shockley Semiconductors included Intel and AMD.

Similarly, Moretti (2012) describes how personal factors led Bill Gates and Paul Allen to move Microsoft from Albuquerque to Seattle, their hometown. At the time, Albuquerque was considered the better place to live, it was favored by most of Microsoft's early employees and the location of many early clients. Yet proximity to family won out, and this decision has reverberated well beyond Microsoft's direct employment. The agglomeration advantages sparked by Microsoft have attracted countless other tech firms to Seattle, including Jeff Bezos relocating from New York City to Seattle when he founded Amazon. Had Gates and Allen not moved home to Seattle, Albuquerque might be home to two of America's three most valued companies in 2020.

A similar and related randomness arises due to the often-serendipitous nature of breakthrough discoveries and their outsized subsequent importance. Zucker, Darby, and Brewer (1998) show that the location of biotech industry follows the positioning of star scientists in the nascent field, and the surging prominence of Toronto for artificial intelligence harkens back to the choice of some key early researchers to locate there, well before the field became so prominent. Duranton (2007) formalizes how random breakthroughs could lead to shifts in the leadership of cities for a tech field or industry, such as the migration of semiconductors from Boston to Silicon Valley. Kerr (2010) quantifies this pattern of reallocation across 36 patenting sectors since the 1970s.

While random sparks play a role, the same breakthroughs often occur contemporaneously in two or more locations (Ganguli, Lin, and Reynolds 2019). Accordingly, a new line of work considers the factors that shape which location emerges as the winner. Duran and Nanda (2019), for example, study the widespread experimentation during the late 1890s and early 1900s as local automobile assemblers learned about the fit between this emerging industry and their city. Despite having fewer entrants initially, activity coalesced in smaller cities—Cleveland, Indianapolis, St. Louis, and Detroit—with Detroit being the ultimate winner by the late 1920s. The smaller city advantage may have been due to the higher physical proximity of relevant stakeholders, allowing for easier experimentation, prototyping, and circulation of ideas. So long as smaller cities had sufficient local input supplies, they may have provided more attention and financial support to the new technology compared to larger markets and fostered relational contracts.

This stream of research yields some tentative conclusions for policymakers. Lerner (2009) documents the poor past performance of public efforts to engineer a cluster from scratch, and Ferrary and Granovetter (2009) blame the widespread failure of policymakers to replicate the success of Silicon Valley on their misunderstanding of complex innovation networks and to the shallowness of venture capital

markets. The unique origin of each existing tech cluster suggests future efforts to seed from scratch are likely to be similarly frustrating.

Instead, a better return is likely to come from efforts to reduce the local costs of experimentation with ideas (Kerr, Nanda, and Rhodes-Kropf 2014), alongside the provision of a good quality of life. There is also likely a role for cities that have developed a position in an emerging sector, even if by random accident due to family ties, to increase the odds they are favored in the shakeout process. Such support is more likely to work if it is broad-based to a sector and avoids attempting to "pick winners" by targeting individual companies. Other cities can take the strategy of increasing their connectivity to leading centers via remote work. Tulsa Remote pays qualified workers with remote jobs $10,000 to move to Tulsa, Oklahoma, and similar programs are popping up elsewhere. Rather than seeking to "become the next Silicon Valley," these efforts focus on connecting with the existing hotspots and being an attractive alternative with a lower cost of living.

Beyond anchor firms, universities also feature prominently in the history of tech clusters, both for the United States and globally (Markusen 1996; Dittmar and Meisenzahl 2020). Under the guidance of Fred Terman, Stanford University fostered a strong relationship with the growing tech community, such as the 1948 creation of the Stanford Industrial Park that would house 11,000 workers from leading tech firms by the 1960s. Famed venture capitalist Arthur Rock summed up the university's driving role around this time: "All of the energetic scientists were forming around Stanford" (as quoted in Lee and Nicholas 2013). Similarly, the placement of a Carnegie-funded library into a city in the decades around 1900 corresponded to a substantial growth in patenting relative to peer cities for the next 20 years (Berkes and Nencka 2019).

Hausman (2012) documents how university innovation fosters local industry growth, and these spillovers can attenuate rapidly (see also Andersson, Quigley, and Wilhelmsson 2009; Kantor and Whalley 2014). With the increase in university patenting following the 1980 Bayh-Dole Act that provided universities greater ownership of intellectual property resulting from government-funded research, these intellectual sparks are growing in number. Universities are also a vibrant source of young, smart workers with frontier skill sets. Marshall (1890) emphasized the benefits of natural advantages like deep harbors and coal mines; strong research universities, along with government-sponsored laboratories, are likely to be key (man-made) natural advantages for new tech clusters. While Silicon Valley was in some ways a blank slate, it did possess from the start a powerful asset with Stanford University.

These historical examples are starting to provide insight that will advance our theory on tech clusters. Duranton and Puga (2001) model a system of cities in which new industries are emerging in large and diverse "nursery" cities. As industries mature and move from experimentation to scale, they no longer value the cross-fertilization enabled by industrial diversity and seek instead to maximize within-sector productivity. The model portrays mature industries as then relocating to less expensive and more specialized cities.

The nursery city model provides a powerful tool for thinking about systems of cities (Henderson 1974). It also fits many industrial experiences, such as the exodus of large-scale apparel manufacturing out of Manhattan over the last century (leaving the Garment District's name and some key fashion designers behind). The nearby "Silicon Alley" in Manhattan's Flatiron district also previously held names like "Toy District" and "Photo District," reflecting the local clusters of previous eras. Yet autos went from cradle to old age in Detroit, and other places like Lowell and Cleveland failed to renew themselves the way New York did. Boston has reinvented itself three times since its colonial days (Glaeser 2005).

What explains these differing fates? One promising hypothesis starts by thinking about the specialization of cities on function versus industry lines (Duranton and Puga 2005). Many models keep industry size much smaller than city size so that reallocation is more likely to happen at the industry level (Duranton and Puga 2001; Duranton 2007). The competitive framework by Porter (1998) emphasizes these radical upheavals that happen at the industry level. By contrast, the historical examples also suggest a fast-growing industry may come to dominate a nursery city so quickly that the city ceases to specialize on a function (like the breeding of new ideas) and instead specializes on an industry (like autos), thereby pushing out the local industry diversity to other locations.[5] The sociology and geography literatures also emphasize local threats to the growth of clusters, such as emerging endogenous barriers to entry (for example, Granovetter 1973).

A richer depiction of these interacting forces connects to many interesting strands of the research literature. Helsley and Strange (2014) model that cities hold a (non-optimal) mix of co-agglomerated industries, due to legacy location choices and persistence. Perhaps a larger city the size of a London or Tokyo is protected from becoming too hyper-specialized around any one fast-growing industry. Other work focuses on superstar cities and power couples seeking dual careers (for example, Gyourko, Mayer, and Sinai 2013; Costa and Kahn 2000). Maybe New York's greatest lever for long-term economic sustainability is that the two members of a high-income couple can have as daring a career as a fintech entrepreneur and as conservative a career as a healthcare top executive, so long as they can also afford to pay $40,000 for their kid's pre-school.

Future Directions for Research

There are many open questions regarding tech clusters, and we conclude with some promising areas of inquiry. Just as tech clusters lead to spillovers across

[5] The spatial equilibrium model also struggles with aspects of the distribution of entrepreneurship across cities (for example, Glaeser 2008; Glaeser, Kerr, and Ponzetto 2010). Recent contributions to the underpinning of a system of cities model include Behrens, Duranton, and Robert-Nicoud (2014) and Davis and Dingel (2019), which provide further references.

technological and industrial boundaries in the real economy, we expect that research on tech clusters will also spill over into and across other fields of economic inquiry.

New employer-employee datasets will allow researchers to quantify the creation and scaling of enterprises inside tech clusters. This step can build upon administrative data, such as the Census Bureau's Longitudinal Employer Household Survey, combined via external links to patenting and venture capital data. Others will take advantage of private datasets like LinkedIn, which is almost a pseudo-Census of the tech industry. For example, these analyses will help differentiate among the many theoretical channels for labor market pooling, ranging from greater matching to insuring workers against the risk of job separation.

Fine-tuned establishment data also facilitate new inquiries. Relatively few studies explore the internal choices within firms for how to locate their many activities, a decision that often involves a tradeoff between proximity to sources of external insight and internal communication and alignment (for recent examples, see Alcácer and Delgado 2016; Lychagin et al. 2016; Kerr 2018). As technology grows in importance, companies appear to be placing more key decision-makers and innovation personnel into tech clusters. Researchers need to develop a better understanding of these location decisions and their global consequences. For example, Landier, Nair, and Wulf (2009) quantify the greater likelihood of business leaders to close plants farther away from the corporate headquarters.

These types of data will further refine our understanding of local economic spillovers in tech clusters. Moretti (2012) calculates that knowledge work creates five non-technical jobs for each knowledge worker, a local multiplier that is substantially higher than manufacturing. These generated jobs also pay better than similar work in other cities. Samila and Sorenson (2011) quantify how venture capital similarly creates new jobs in local areas beyond the start-ups directly supported, and that these tend to be well-paid positions, but that the magnitude is overall modest in nature. The resulting escalation of real estate rents, however, also crowds out lower income individuals (Gyourko, Mayer, and Sinai 2013), and a more complete portrait of the benefits and strains for local areas from blossoming tech work is needed.

Emerging research is also exploring how tech clusters shape the careers of individuals and the early stages of companies. Gompers, Lerner, and Scharfstein (2005) document how many venture capital-backed entrepreneurs cut their teeth through prior work in startups, and Moretti (2019) estimates that inventors moving to a larger tech cluster experience increases in their patenting outcomes. Future work can extend this person-level perspective to see how cities shape the types of work created by inventors. In a similar way, Guzman (2019) documents the migration of startups from their founding city to Silicon Valley. Higher-quality firms are more likely to migrate to Silicon Valley, where they appear to receive better knowledge spillovers.

Will the existing tech clusters strengthen going forward? A simple extrapolation of trend lines suggests greater spatial concentration for tech clusters looms on the horizon. Indeed, many policy proposals—ranging from pushing massive

stimulus of basic research and development spending into the heartland (Gruber and Johnson 2019) to creating regionally capped visa allocations for skilled immigrants—start with the premise that, because tech clusters are becoming more concentrated, policymakers need to step in. Due to lower agglomeration benefits outside of tech clusters, these proposals to push activity into other cities and regions are typically based upon achieving regional equity and political buy-in and may face a possible tradeoff of reduced aggregate economic output. Moretti (2019) estimates, for example, that the special concentration of inventors into leading tech centers boosts innovation by 11 percent, compared to a scenario where all inventors spread out evenly over cities. Additional research to quantify the particular role of tech clusters and their innovations (both in total number and their traits like atypical combinations) into economic growth will be very valuable.

Yet many factors may naturally limit further spatial inequality. Doubling Silicon Valley's size—which is impossible on many geographic and political levels—would still only make it 2 percent of the US population. We are witnessing a major transformation of business to achieve appropriate positions in powerful tech hubs, but most workers and consumers will always be far away. Large companies will only pay the hefty prices of tech clusters for some key workers, instead investing to ensure that the firm transmits the important information effectively to others in the company. At the local level, political pressures to limit housing construction will make it costly for certain tech centers to expand: for example, Hsieh and Moretti (2019) estimate that housing constraints that limited the spatial reallocation of workers towards the most productive cities of New York and the San Francisco Bay area lowered US growth by 36 percent since the 1960s. Political tensions and spatial disparities across US regions may also limit how big tech clusters can become.

These factors were already in play in early 2020 when the COVID-19 crisis added yet more complexity to the future of tech clusters. On one hand, the acceleration in technology adoption brought about by the pandemic—for example, to shift activity towards e-commerce or contactless stores—is likely to increase the near-term importance of tech clusters. Efforts by tech companies to provide assistance in the crisis have also helped repair some of the reputation hits they recently incurred. Yet tech clusters have thrived on physical proximity, which can unfortunately transmit viruses as easily as ideas as well as on global talent and trade. These benefits may be dampened in years ahead due to the virus itself, along with the follow-on business and political changes it produces. Catalysts like venture capital funding may also be in shorter supply in years ahead. The man-made nature of tech clusters leaves them more malleable than those built around harbors or coal mines, and future research will shed more light on tech clusters through the adjustments that lie ahead.

■ *The authors thank Harald Bathelt, Neil Coe, Ed Glaeser, Gordon Hanson, Enrico Moretti, Ramana Nanda, Will Strange, Timothy Taylor, and Heidi Williams for their insightful thoughts, comments, or feedback on this paper. The authors also thank Brad Chattergoon, Maggie Dalton, Brad DeSanctis, and Louis Maiden for excellent research assistance.*

References

Acs, Zoltan J., Luc Anselin, and Attila Varga. 2002. "Patents and Innovation Counts as Measures of Regional Production of New Knowledge." *Research Policy* 31 (7): 1069–85.

Agrawal, Ajay, and Iain Cockburn. 2003. "The Anchor Tenant Hypothesis: Exploring the Role of Large, Local, R&D-intensive Firms in Regional Innovation Systems." *International Journal of Industrial Organization* 21 (9): 1217–53.

Agrawal, Ajay, Iain Cockburn, Alberto Galasso, and Alexander Oettl. 2014. "Why Are Some Regions More Innovative Than Others? The Role of Small Firms in the Presence of Large Labs." *Journal of Urban Economics* 81 (1): 149–65.

Agrawal, Ajay, Iain Cockburn, and Carlos Rosell. 2010. "Not Invented Here? Innovation in Company Towns." *Journal of Urban Economics* 67 (1): 78–89.

Alcácer, Juan, and Wilbur Chung. 2007. "Location Strategies and Knowledge Spillovers." *Management Science* 53 (5): 760–76.

Alcácer Juan, and Wilbur Chung. 2014. "Location Strategies for Agglomeration Economies." *Strategic Management Journal* 35 (12): 1749–61.

Alcácer, Juan, and Mercedes Delgado. 2016. "Spatial Organization of Firms and Location Choices Through the Value Chain." *Management Science* 62 (11): 3213–34.

Almazan, Andres, Adolfo De Motta, and Sheridan Titman. 2007. "Firm Location and the Creation and Utilization of Human Capital." *Review of Economic Studies* 74 (4): 1305–27.

Almeida, Paul, and Bruce Kogut. 1999. "Localization of Knowledge and the Mobility of Engineers in Regional Networks." *Management Science* 45 (7): 905–17.

Andersson, Roland E., John M. Quigley, and Mats Wilhelmsson. 2009. "Higher Education, Localization and Innovation: Evidence from a Natural Experiment." *Journal of Urban Economics* 66 (1): 2–15.

Arzaghi, Mohammad, and J. Vernon Henderson. 2008. "Networking off Madison Avenue." *Review of Economic Studies* 75 (4): 1011–38.

Audretsch, David B., and Maryann P. Feldman. 1996. "R&D Spillovers and the Geography of Innovation and Production." *American Economic Review* 86 (3): 630–40.

Balachandran, Sarath, and Exequiel Hernandez. 2019. "Mi Casa Es Tu Casa: Immigrant Entrepreneurs as Pathways to Foreign Venture Capital Investments." http://dx.doi.org/10.2139/ssrn.3331264.

Baldwin, Carliss Y., and Clark, Kim B. 1997. "Managing in an Age of Modularity." *Harvard Business Review* (September-October).

Bathelt, Harald, and Peng-Fei Li. 2014. "Global Cluster Networks—Foreign Direct Investment Flows from Canada to China." *Journal of Economic Geography* 14 (1): 45–71.

Bathelt, Harald, Anders Malmberg, and Peter Maskell. 2004. "Clusters and Knowledge: Local Buzz, Global Pipelines and the Process of Knowledge Creation." *Progress in Human Geography* 28 (1): 31–56.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. "Productive Cities: Sorting, Selection, and Agglomeration." *Journal of Political Economy* 122 (3): 507–53.

Berger, Suzanne. 2005. *How We Compete: What Companies Around the World Are Doing to Make It in Today's Global Economy*. New York: Doubleday.

Bergquist, Kyle, Carsten Fink, and Julio Raffo. 2017. "Identifying and Ranking the World's Largest Clusters of Inventive Activity." WIPO Working Paper 34.

Berkes, Enrico, and Ruben Gaetani. 2019. "The Geography of Unconventional Innovation." Rotman School of Management Working Paper 3423143.

Berkes, Enrico, and Peter Nencka. 2019. "'Novel' Ideas: The Effects of Carnegie Libraries on Innovative Activities." Unpublished.

Branstetter, Lee, Guangwei Li, and Francisco Veloso. 2015. "The Rise of International Coinvention." In: Jaffe, A., Jones, B. (Eds.) *The Changing Frontier: Rethinking Science and Innovation Policy*, University of Chicago Press 135–68.

Breschi, Stefano, and Francesco Lissoni. 2009. "Mobility of Skilled Workers and Co-invention Networks: An Anatomy of Localized Knowledge Flows." *Journal of Economic Geography* 9 (4), 439–68.

Bresnahan, Timothy, and Alfonso Gambardella, eds. 2001. *Building High-Tech Clusters: Silicon Valley and Beyond*. Cambridge: Cambridge University Press.

Carlino, Gerald A., Jake K. Carr, Robert M. Hunt, and Tony E. Smith. 2012. "The Agglomeration of R&D Labs." Federal Reserve Bank of Philadelphia Working Paper 12–22.

Carlino, Gerald A., Satyajit Chatterjee, and Robert M. Hunt. 2007. "Urban Density and the Rate of Invention." *Journal of Urban Economics* 61 (3): 389–419.

Carlino, Gerald, and William Kerr. 2015. "Agglomeration and Innovation." In *Handbook of Regional and Urban Economics,* Volume 5, edited by Gilles Duranton, J. Vernon Henderson, and William Strange, 349–404. Amsterdam: Elsevier.

Christophe Carrincazeaux, Yannick Lung, and Allain Rallet. 2001. "Proximity and Localisation of Corporate R&D Activities." *Research Policy* 30 (5): 777–89.

Chatterji, Aaron, Edward Glaeser, and William Kerr. 2014. "Clusters of Entrepreneurship and Innovation." In *Innovation Policy and the Economy 2013*, Vol. 14, edited by Josh Lerner and Scott Stern, 129–66. Chicago: University of Chicago Press.

Choudhury, Prithwiraj. 2017. "Innovation Outcomes in a Distributed Organization: Intrafirm Mobility and Access to Resources." *Organization Science* 28 (2): 339–54.

Choudhury, Prithwiraj. 2016. "Return Migration and Geography of Innovation in MNEs: A Natural Experiment of Knowledge Production by Local Workers Reporting to Return Migrants." *Journal of Economic Geography*, 16 (3): 585–610.

Coe, Neil M., and Timothy G. Bunnell. 2003. "'Spatializing' Knowledge Communities: Towards a Conceptualization of Transnational Innovation Networks." *Global Networks* 3 (4): 437–56.

Combes, Pierre Philippe, and Gilles Duranton. 2006. "Labour Pooling, Labour Poaching and Spatial Clustering." *Regional Science and Urban Economics* 36 (1): 1–28.

Costa, Dora L., and Matthew E. Kahn. 2000. "Power Couples: Changes in the Locational Choice of the College Educated, 1940–1990." *Quarterly Journal of Economics* 115 (4): 1287–1315.

Davis, Donald R., and Jonathan I. Dingel. 2019. "A Spatial Knowledge Economy." *American Economic Review* 109 (1): 153–70.

Delgado, Mercedes, Michael E. Porter, and Scott Stern. 2010. "Clusters and Entrepreneurship." *Journal of Economic Geography* 10 (4): 495–518.

Dittmar, Jeremiah, and Ralf Meisenzahl. 2020. "Public Goods Institutions, Human Capital, and Growth: Evidence from German History." *The Review of Economic Studies* 87 (2): 959–96.

Duran, Xavier, and Ramana Nanda. 2019. "Experimentation in the Early U.S. Automobile Industry." Unpublished.

Duranton, Gilles. 2007. "Urban Evolutions: The Fast, the Slow, and the Still." *American Economic Review* 97 (1): 197–221.

Duranton, Gilles. 2011. "California Dreamin': The Feeble Case for Cluster Policies." *Review of Economic Analysis* 3: 3–45.

Duranton, Gilles, and Diego Puga. 2001. "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products." *American Economic Review* 91 (5): 1454–77.

Duranton, Gilles, and Diego Puga. 2004. "Micro-foundations of Urban Agglomeration Economies." In *Handbook of Urban and Regional Economics*, Vol. 4, edited by J. Vernon Henderson and J. F. Thisse, chapter 48. Amsterdam: Elsevier.

Duranton, Gilles, and Diego Puga. 2005. "From Sectoral to Functional Urban Specialisation." *Journal of Urban Economics* 57 (2): 343–70.

Egan, Edward J., Anne Dayton, Diana Carranza. 2017. *The Top 100 U.S. Startup Cities in 2016.* Houston, TX: Rice University's Baker Institute for Public Policy.

Ellison, Glenn, and Edward L. Glaeser. 1997. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach." *Journal of Political Economy* 105 (5): 889–927.

Ellison, Glenn, Edward Glaeser, and William Kerr. 2010. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review* 100 (3): 1195–1213.

Faggio, Giulia, Olmo Silva, and William C. Strange. 2017. "Heterogeneous Agglomeration." *Review of Economics and Statistics* 99 (1): 80–94.

Fairlee, Robert W., and Aaron K. Chatterji. 2013. "High-Technology Entrepreneurship in Silicon Valley." *Journal of Economics and Management Strategy* 22 (2): 365–89.

Falck, Oliver, Michael Fritsch, and Stephan Heblich. 2011. "The Phantom of the Opera: Cultural Amenities, Human Capital, and Regional Economic Growth." *Labour Economics* 18 (6): 755–66.

Fallick, Bruce, Charles A. Fleischman, and James B. Rebitzer. 2006. "Job-Hopping in Silicon Valley: Some Evidence Concerning the Microfoundations of a High-Technology Cluster." *Review of Economics and Statistics* 88 (3): 472–81.

Feldman, Maryann. 2003. "The Locational Dynamics of the US Biotech Industry: Knowledge Externalities and the Anchor Hypothesis." *Industry and Innovation* 10 (3): 311–29.

Feldman, Maryann, and Dieter Kogler. 2010. "Stylized Facts in the Geography of Innovation." In *Handbook of the Economics of Innovation*, Vol. 1, edited by Bronwyn Hall and Nathan Rosenberg, 381–410. Amsterdam: Elsevier.

Ferrary, Michael, and Mark Granovetter. 2009. "The Role of Venture Capital Firms in Silicon Valley's Complex Innovation Network." *Economy and Society* 38 (2): 326–59.

Fleming, Lee, and Matt Marx. 2006. "Managing Creativity in Small Worlds." *California Management Review* 48 (4): 6–27.

Florida, Richard. 2005. *Cities and the Creative Class.* Routledge: New York.

Gagne, J. F. 2019. "Global AI Talent Report 2019." jfgagne.ai/talent-2019.

Ganguli, Ina, Jeffrey Lin, and Nicholas Reynolds. 2020. "The Paper Trail of Knowledge Spillovers: Evidence from Patent Interferences." *American Economic Journal: Applied Economics* 12 (2): 278–302.

Gerlach, Heiko, Thomas Rønde, and Konrad Stahl. 2009. "Labor Pooling in R&D Intensive Industries." *Journal of Urban Economics* 65 (1): 99–111.

Gilson, Ronald J. 1999. "The Legal Infrastructure of High Technology Industrial Districts: Silicon Valley, Route 128, and Covenants Not to Compete." *New York University Law Review* 74 (3): 575–629.

Glaeser, Edward. 2008. *Cities, Agglomeration and Spatial Equilibrium.* Oxford: Oxford University Press.

Glaeser, Edward. 2005. "Reinventing Boston: 1630–2003." *Journal of Economic Geography* 5 (2): 119–53.

Glaeser, Edward, Heidi D. Kallal, José A. Scheinkman, and Andrew Shleifer. 1992. "Growth in Cities." *Journal of Political Economy* 100 (6): 1126–52.

Glaeser, Edward, and William Kerr. 2009. "Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?" *Journal of Economics and Management Strategy* 18 (3): 623–63.

Glaeser, Edward, Sari Pekkala Kerr, and William Kerr. 2015. "Entrepreneurship and Urban Growth: An Empirical Assessment with Historical Mines." *Review of Economics and Statistics* 97 (2): 498–520.

Glaeser, Edward, William Kerr, and Giacomo A. M. Ponzetto. 2010. "Clusters of Entrepreneurship." *Journal of Urban Economics* 67 (1): 150–68.

Gompers, Paul, Josh Lerner, and David Scharfstein. 2005. "Entrepreneurial Spawning: Public Corporations and the Genesis of New Ventures, 1986 to 1999." *Journal of Finance* 60 (2): 577–614.

Gompers, Paul, and Sophie Q. Wang. 2017. "Diversity in Innovation." Harvard Business School Working Paper 17-067.

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360–80.

Groysberg, Boris. 2010. *Chasing Stars: The Myth of Talent and the Portability of Performance.* Princeton: Princeton University Press.

Gruber, Jonathan, and Simon Johnson. 2019. *Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream.* New York: PublicAffairs.

Guzman, Jorge, and Scott Stern. 2019. "The State of American Entrepreneurship: New Estimates of the Quality and Quantity of Entrepreneurship for 32 US states, 1988–2014." NBER Working Paper 22095.

Gyourko, Joseph, Christopher Mayer, and Todd Sinai. 2013. "Superstar Cities." *American Economic Journal: Economic Policy* 5 (4): 167–99.

Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg. 2001. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." NBER Working Paper 8498.

Hausman, Naomi. 2012. "University Innovation, Local Economic Growth, and Entrepreneurship." US Census Bureau Center for Economic Studies Paper CES-WP-12-10.

Hausman, Naomi. 2019. "Non-compete Law, Labor Mobility, and Entrepreneurship." Unpublished.

Hellmann, Thomas, and Enrico Perotti, 2011. "The Circulation of Ideas in Firms and Markets." *Management Science*, 57 (10): 1813–26.

Helsley, Robert, and William Strange. 1990. "Matching and Agglomeration Economies in a System of Cities." *Regional Science and Urban Economics*, 20 (2): 189–212.

Helsley, Robert, and William Strange. 2002. "Innovation and Input Sharing." *Journal of Urban Economics* 51 (1): 25–45.

Helsley, Robert, and William Strange. 2014. "Coagglomeration, Clusters, and the Scale and Composition of Cities." *Journal of Political Economy* 122 (5): 1064–93.

Henderson, J. V. 1974. "The Sizes and Types of Cities." *American Economic Review* 64 (4): 640–56.

Henderson, Vernon, Ari Kuncoro, and Matt Turner. 1995. "Industrial Development in Cities." *Journal of Political Economy* 103 (5): 1067–90.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation" *American*

Economic Journal: Macroeconomics 11 (2): 1–39.
Humphrey, John, and Hubert Schmitz. 2002. "How Does Insertion in Global Value Chains Affect Upgrading in Industrial Clusters?" *Regional Studies* 36 (9): 1017–1102.
Jacobs, Jane. 1970. *The Economy of Cities*. New York: Vintage Books.
Jaffe, Adam, Manuel Trajtenberg, and Michael Fogarty. 2000. "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors." *American Economic Review* 90 (2): 215–18.
Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics* 108 (3): 577–98.
Kantor, Shawn, and Alexander Whalley. 2014. "Knowledge Spillovers from Research Universities: Evidence from Endowment Value Shocks." *Review of Economics and Statistics* 96 (1): 171–88.
Kenny, Martin. 2000. *Understanding Silicon Valley: The Anatomy of an Entrepreneurial Region*. Stanford: Stanford University Press.
Kerr, Sari Pekkala, and William Kerr. 2018. "Global Collaborative Patents." *The Economic Journal* 128: F235–72.
Kerr, Sari Pekkala, and William Kerr. 2020. "Immigrant Entrepreneurship in America: Evidence from the Survey of Business Owners 2007 & 2012." *Research Policy* 49 (3): 103918.
Kerr, William. 2010. "Breakthrough Inventions and Migrating Clusters of Innovation." *Journal of Urban Economics* 67 (1): 46–60.
Kerr, William. 2018. "Navigating Talent Hot Spots." *Harvard Business Review*. 95 (5): 80–86
Kerr, William. 2019. *The Gift of Global Talent: How Migration Shapes Business, Economy, and Society* . Stanford: Stanford University Press.
Kerr, William, and Scott Duke Kominers. 2015. "Agglomerative Forces and Cluster Shapes." *Review of Economics and Statistics* 97 (4): 877–99.
Kerr, William, and William Lincoln. 2010. "The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention." *Journal of Labor Economics* 28 (3): 473–508.
Kerr, William, Ramana Nanda, and Matthew Rhodes-Kropf. 2014. "Entrepreneurship as Experimentation." *Journal of Economic Perspectives* 28 (3): 25–48.
Klepper, Steven. 2010. "The Origin and Growth of Industry Clusters: The Making of Silicon Valley and Detroit." *Journal of Urban Economics* 67 (1): 15–32.
Klepper, Steven. 2016. *Experimental Capitalism: The Nanoeconomics of American High-Tech Industries*. Princeton: Princeton University Press.
Krugman, Paul. 1991. *Geography and Trade*. Cambridge, MA: MIT Press.
Lamoreaux, Naomi, Margaret Levenstein, and Kenneth Sokoloff. 2004. "Financing Invention during the Second Industrial Revolution: Cleveland, Ohio, 1870–1920." NBER Working Paper 10923.
Landier, Augustin, Vinay Nair, and Julie Wulf. 2009. "Trade-offs in Staying Close: Corporate Decision Making and Geographic Dispersion." *Review of Financial Studies* 22 (3): 1119–48.
Lee, James, and Tom Nicholas. 2013. "The Origins and Development of Silicon Valley." Harvard Business School Case 813-098.
Lerner, Josh. 2009. *Boulevard of Broken Dreams: Why Public Efforts to Boost Entrepreneurship and Venture Capital Have Failed-and What to Do about It*. Princeton: Princeton University Press.
Liang, Wen-Jung, Chao-Cheng Mai, Jacques-François Thisse, and Ping Wang. 2019. "On the Economics of Science Parks." NBER Working Paper 25595.
Lin, Jeffrey. 2011. "Technological Adaptation, Cities, and New Work." *Review of Economics and Statistics* 93 (2): 554–74.
Lychagin, Sergey, Joris Pinkse, Margaret Slade, and John Van Reenen. 2016. "Spillovers in Space: Does Geography Matter?" *Journal of Industrial Economics* 64 (2): 295–335.
Malmberg, Anders, and Dominic Power. 2005. "(How) Do (Firms in) Clusters Create Knowledge?" *Industry and Innovation* 12 (4): 409–31.
Markusen, Ann. 1996. "Sticky Places in Slippery Space: A Typology of Industrial Districts." *Economic Geography* 72 (3): 293–313.
Marshall, Aarian. 2018. "Uber and Waymo Abruptly Settle For $245 Million." *Wired*, February 9. https://www.wired.com/story/uber-waymo-lawsuit-settlement.
Marshall, Alfred. 1890. *Principles of Economics*. London: Macmillan.
Marx, Matt, Deborah Strumsky, and Lee Fleming. 2009. "Mobility, Skills, and the Michigan Non-Compete Experiment." *Management Science* 55 (6): 875–89.
Marx, Matt, Jasjit Singh, and Lee Fleming. 2015. "Regional Disadvantage? Employee Non-Compete Agreements and Brain Drain." *Research Policy* 44 (2): 394–404.

Matouschek, Niko, and Frédéric Robert-Nicoud. 2005. "The Role of Human Capital Investments in the Location Decision of Firms." *Regional Science and Urban Economics* 35 (5): 570–83.

Mehrotra, Kartikay. 2016. "Samsung, LG Accused of Silicon Valley Anti-Poaching Deal," Bloomberg, September 12, https://www.bloomberg.com/news/articles/2016-09-12/samsung-lg-accused-of-silicon-valley-anti-poaching-agreement.

Miguelez, Ernest. 2014. "Inventor Diasporas and the Internationalization of Technology." Discussion Paper 25/14, Centre for Research & Analysis of Migration.

Moretti, Enrico. 2012. *The New Geography of Jobs*. New York: Houghton Mifflin Harcourt.

Moretti, Enrico. 2019. "The Effect of High-Tech Clusters on the Productivity of Top Inventors." NBER Working Paper 26270.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura. 2014. "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach." *Review of Economics and Statistics* 96 (5): 967–85.

Nanda, Ramana, and Tarun Khanna. 2010. "Diasporas and Domestic Entrepreneurs: Evidence from the Indian Software Industry." *Journal of Economics and Management Strategy* 19 (4): 991–1012.

Nathan, Max. 2014. "The Wider Economic Impacts of High-Skilled Migrants: A Survey of the Literature for Receiving Countries." *IZA Journal of Migration* 3 (4).

Nathan, Max. 2015. "Same Difference? Minority Ethnic Inventors, Diversity and Innovation in the UK." *Journal of Economic Geography* 15 (1): 129–68.

National Science Foundation. 2020. National Center for Science and Engineering Statistics | NSF 20-311. https://ncses.nsf.gov/pubs/nsf20311. (accessed June 17, 2020).

Olson, Gary, and Judith Olson. 2003. "Mitigating the Effects of Distance on Collaborative Intellectual Work." *Economics of Innovation and New Technology* 12 (1): 27–42.

Overman, Henry G., and Diego Puga. 2010. "Labour Pooling as a Source of Agglomeration: An Empirical Investigation." In *Agglomeration Economics,* edited by Edward Glaeser, 133–50. Chicago: University of Chicago Press.

Peri, Giovanni, Kevin Shih, and Chad Sparber. 2015. "STEM Workers, H-1B Visas and Productivity in U.S. cities." *Journal of Labor Economics* 33 (S1): S225–55.

Porter, Michael E. 1998. "Clusters and the New Economics of Competition." *Harvard Business Review.*

Roberts, Jeff John. 2015. "Tech Workers Will Get Average of $5,770 under Final Anti-Poaching Settlement." *Fortune,* September 3. http://fortune.com/2015/09/03/koh-anti-poach-order.

Rosenthal, Stuart, and William Strange. 2001. "The Determinants of Agglomeration." *Journal of Urban Economics* 50: 191–229.

Rosenthal, Stuart, and William Strange. 2003. "Geography, Industrial Organization, and Agglomeration." *Review of Economics and Statistics* 85 (2): 377–93.

Rotemberg, Julio, and Garth Saloner. 2000. "Competition and Human Capital Accumulation: A Theory of Interregional Specialization and Trade." *Regional Science and Urban Economics* 30 (4): 373–404.

Samila, Sampsa, and Olay Sorenson. 2011. "Venture Capital, Entrepreneurship and Economic Growth." *Review of Economics and Statistics* 93 (1): 338–49.

Saxenian, AnnaLee. 1991. "The Origins and Dynamics of Production Networks in Silicon Valley." *Research Policy* 20 (5): 423–37.

Saxenian, AnnaLee. 1994. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Saxenian, AnnaLee. 2007. *The New Argonauts: Regional Advantage in a Global Economy*. Cambridge, MA: Harvard University Press.

Saxenian, AnnaLee, Yasuyuki Motoyama, and Xiaohong Quan. 2002. *Local and Global Networks of Immigrant Professionals in Silicon Valley*. San Francisco: Public Policy Institute of California.

Singer, Stephen. 2016. "General Electric CEO lauds Boston as new site." *Hartford Courant*, March 24. https://www.courant.com/business/hc-immelt-ge-boston-connecticut-20160324-story.html.

Singh, Jasjit, and Matt Marx. 2013. "Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity." *Management Science* 59 (9): 2056–78.

Storper, Michael, and Anthony J. Venables. 2004. "Buzz: Face-to-Face Contact and the Urban Economy." *Journal of Economic Geography* 4 (4): 351–70.

Sturgeon, Timothy J. 2002. "Modular Production Networks: A New American Model of Industrial Organization." *Industrial and Corporate Change* 11 (3): 451–96.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 342 (6157): 468–72.

Waldinger, Fabian. 2012. "Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi German." *The Review of Economic Studies* 79 (2): 838–61.

World Intellectual Property Organization. 2017. *The Global Innovation Index 2017: Innovation Feeding the World*. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2017.pdf. (accessed june 17, 2020).

Zucker, Lynne, Michael Darby, and Marilynn Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review* 88 (1): 290–306.

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 77–98

Internal Mobility: The Greater Responsiveness of Foreign-Born to Economic Conditions

Gaetano Basso and Giovanni Peri

Geographic mobility is an important mechanism in determining the spatial distribution of economic activity and local economic growth. From this perspective, mobility happens in response to employment opportunities and differentials in real income (that is, in local earnings relative to local prices). Not all mobility in the United States is driven by economic motives. About 30 percent of the residential mobility is attributed to "family" reasons, such as establishing a new household or changing family status, and about 5 percent to other reasons related to health, climate, and education (Ihrke 2014). Nevertheless, mobility for economic reasons, usually related to jobs and housing, is the stated rationale for the majority of moves. Moreover, this type of mobility plays an important role in spatial convergence of local real income as well as in determining a spatial equilibrium—with population growth occurring where demand for labor grows and population decline occurring where labor demand decreases (Rosen 1979; Roback 1982). It has become a point of concern that the process of regional economic convergence in the US economy seems to have slowed down in the 1980s (Ganong and Shoag 2017), and the economic fortunes of US cities started diverging from each other in wages, employment growth, and productivity (Moretti 2012). Meanwhile, internal geographic mobility seems to be on a declining trend, as well (Molloy, Smith, and Wozniak 2011). While the patterns of regional and urban convergence and divergence are driven by several factors, some local and other global, migration does not seem to be as much of a partially counterbalancing force as it was in the past.

■ *Gaetano Basso is an Economic Adviser, Bank of Italy, Rome, Italy. Giovanni Peri is Professor of Economics, University of California, Davis, Davis, California. Their email addresses are gaetano.basso@bancaditalia.it and gperi@ucdavis.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.77.

Our goal in this article is to provide a review of internal mobility in the United States in recent decades, with a focus on the period since 2000 and an emphasis on economically motivated mobility. We first extend and update data on total mobility across US states and labor markets, beginning in 1980, with a focus on the recent period, 2000–2017. We confirm the pattern previously established in this literature of an ongoing decline in mobility, using a combination of data from the American Community Survey, the decennial Census, and the Internal Revenue Service. We then focus on foreign-born individuals and establish that, on average, they do not have total mobility rates higher than natives; in fact, they are somewhat smaller. However, we also identify a dimension over which foreign-born mobility varies substantially: the newly arrived foreign-born with less than ten years in the United States are much more mobile across states and labor markets than natives.

The foreign-born population group is contributing in an important way to the evolution of the spatial distribution of labor in the United States, if for no other reason than because 43 percent of US labor force growth since 2000 has been due to immigrants. But when we dig more deeply, we also observe that geographic mobility of the foreign-born in response to local employment shocks is higher than their proportion in the population. Indeed, during the period 1980–2000, when inflow of new immigrants was large, the foreign-born population responded much more strongly than did the native population to differential employment growth across labor markets. As a consequence, highly successful cities became cities with higher immigrant density by the year 2000. In the period 2000–2017, which includes a deep recession and strong recovery and during which new migration from abroad declined and the long-term immigrants became a more sizeable group, the foreign-born population still responded more than proportionally to local growth in labor demand. This was mainly because cities with large immigrant shares performed better than those with small shares of immigrants, and although the long-term immigrants in the United States were not very mobile, the network effects of previous immigrants implied the continued settlement of new immigrants in those cities.

We review potential explanations for the disproportionate role of foreign-born individuals in the population response to local increases in labor demand—ranging from differential exposure to housing and local prices, to the role of early enclaves and persistent demand shocks, and to their distribution and specialization across occupations. Each of these explanations contributes to our understanding of the special role of foreign-born individuals and their greater propensity to respond to growth in labor demand. We will also suggest some promising and less-explored avenues to understand this phenomenon more fully.

Measures of Total Mobility

To connect with the previous literature on US internal mobility and specifically with patterns presented in this journal by Molloy, Smith, and Wozniak (2011) who provided a consistent series capturing total internal mobility in the United

States from 1980 to 2008, we begin the paper by updating information on interstate and inter-labor market mobility. In particular, we demonstrate that the decline in mobility apparent in earlier work has continued since the Great Recession.

Mobility across States

Figure 1 shows the trends in annual total migration rates across states calculated as the percentage ratio between in-migrants and the resident population in a given year: that is, we focus on people who change their residence between two consecutive years. This measure is the object of analysis in several recent studies of internal mobility in the US economy (Albert and Monras 2018; Amior 2019; Ganong and Shoag 2017; Kaplan and Schulhofer-Wohl 2017).

We rely on three sources for the figure. As a starting point, the American Community Survey has asked directly about annual mobility since the first (experimental) version of the survey was fielded in 2000. The decennial Censuses of 1980, 1990, and 2000 ask about five-year mobility rates, not one-year rates. These questions allow information on origin-destination pairs, at least for US states, and cover the entire US population. Large samples of both decennial Census and American Community Survey data are freely available to researchers through IPUMS (Ruggles et al. 2019). To derive one-year mobility rates using the decennial Census data from 1980 and 1990, we use the ratio between five-year and one-year mobility rates in 2000, a year when the Census Bureau ran both the decennial survey and the first American Community Survey.[1]

Using this data from the decennial Census and the American Community Survey is better than relying on other data sources that have been used to study mobility, like the Current Population Survey (CPS) and its Annual Social and Economic Supplement, for two primary reasons.[2] First, Census and American Community Survey samples are significantly larger than any other data source

[1] The comparison of the 2000 five-year rates from Census and of the one-year rate from the 2000 American Community Survey shows that some people likely move more than once in five years, returning to their home states (or that they have recollection bias in how they answer the survey). In fact, the ratio of the quinquennial to annual migration rates is about 0.71 and is stable across demographic groups. Kaplan and Schulhofer-Wohl (2017), among others, suggest not using the 2000-2004 American Community Survey data, because the sample is smaller and the program was experimental. However, for purposes of documenting the long-term trends in state migration rates, the 2000-2004 American Community Surveys appear reliable. Further details are available in the document accompanying the replication data and programs at the online Appendix available with this article at the *Journal of Economic Perspectives* website.

[2] The Current Population Survey (CPS), and in particular the Annual Social and Economic Supplement (ASEC), while used in several studies, seems less reliable. First, Census and American Community Survey data are based on much larger samples and are cross-sectional in nature, which reduces measurement error and attrition; for example, American Community Survey data cover 1 percent of the US population each year, interviewing about 300,000 thousand participants each month, versus the 60,000 monthly respondents of the CPS. Second, since 1996 the CPS overestimates the decline in interstate migration (due to attrition and imputation) and reports lower levels of internal mobility than other sources (Kaplan and Schulhofer-Wohl 2012). Finally, the CPS has been shown to misreport receipts of government transfer programs (Meyer and Mittag 2019), earnings, and poverty status (Meyer et al. 2019) at an increasing rate over time (Meyer, Mok, and Sullivan 2015): similar concerns may be valid for self-reported migration too.

Figure 1
Interstate Annual Mobility, 1980–2017

Note: The graph represents the population moving across state borders each year as a percentage of the residents in the state of destination. The black dots and the black solid line are constructed using data from decennial Census (1980–1990) and American Community Survey (2000–2017) data, relative to all resident population (excluding those residing in group quarters) and based on the information provided by individuals about their state of residence in the previous and current year. The decennial Census figures have been adjusted from five-year to one-year migration rates according to a correction factor based on the 2000 Census and American Community Survey data, as described in the main text. The grey dashed line is constructed using data from the Internal Revenue Service Statistics of Income (IRS-SOI) calculating the number of tax exemptions that move across state lines in the previous year relative to total tax exemptions' population. The series stops in 2009 because of migration misreporting acknowledged by the IRS-SOI.

available, allowing researchers to study migration patterns for small geographical units such as labor markets, which would be measured with a very significant amount of noise using a much smaller dataset such as the Current Population Survey. Second, the publicly available micro-data from the Census Bureau provide information along many different demographic, economic, and geographic dimensions.

To validate and corroborate the patterns that we identify with the Census data, in Figure 1 we also take advantage of the Internal Revenue Service (IRS) migration statistics, which are freely available from the Statistics of Income website.[3] IRS data

[3]The IRS Statistics of Income website is https://www.irs.gov/statistics/soi-tax-stats-migration-data. The IRS data available on the website cover the period 1990–2018, of which we use 1990–2009. The additional migration data covering the period 1983–1989 were obtained from the IRS. We thank Andrew Foote for his help in locating and using these data.

derive the migration flows from administrative sources and record the change of residence between pairs of counties and pairs of states as reported in the income tax forms filed by residents in two consecutive years. These data only cover tax filers who constitute about 87 percent of US population between 1992 and 2009 (as reported by Molloy, Smith, and Wozniak 2011) and could miss some movers, but they are very consistent over time. It seems plausible that movers in the population of filers likely track movers in the overall population.[4] For Figure 1, the IRS data series is constructed using the total number of tax exemptions who change state in a year as a percentage share of the non-migrating population of the destination state. The IRS Statistics of Income group has acknowledged reliability issues and the existence of migration misreporting from 2010 onwards, which seems to be based on change in the criteria for data collection. As a result, the IRS internal migration data after 2009 are not consistent with the rest of the series, and in Figure 1, we drop those years.

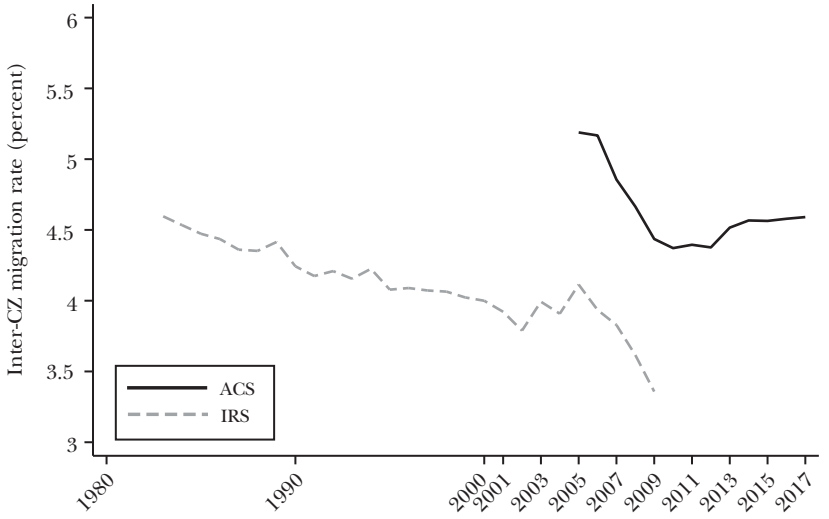
Figure 1 shows two main facts. First, it confirms the long-term decline in interstate migration previously documented by several studies starting with Molloy, Smith, and Wozniak(2011). Between 1980 and 2017, the drop of the migration rate has been around 0.66 percentage points (or totaling about 24 percent of the 1980 value). The decline shown is consistent with what was found already in previous calculations based on similar data. It is not as pronounced, however, as what other studies based on data from the Current Population Survey have found (for example, see Figure 2 in Molloy, Smith, and Wozniak 2011). Second, the IRS and the Census migration-rate series track each other very closely. The gap between the two lines probably exists because the IRS data exclude those who do not file income taxes, who are also known to be somewhat less mobile than tax filers (Molloy, Smith, and Wozniak 2011). However, the difference is small and has not changed much over time. The close correspondence in trends and fluctuations of mobility measures using these two data sources suggests that they are capturing actual trends in the mobility of US citizens. We also notice some fluctuations in yearly mobility post-2000, which seems particularly accentuated in the IRS data and is roughly procyclical, with an increase in mobility pre-2006, a decline 2006–2010, and a recovery in mobility rates after that. These cycles, however, do not seem particularly prominent—especially in the American Community Survey data. The slow but constant decline in state-level mobility over the long run still seems to be the predominant feature of the data.

Mobility across Commuting Zones

Making connections from state-level mobility to economic variables can be tricky. Some urban areas sit astride a state boundary, so it is possible to move one's residence across a state boundary while remaining in the same local labor

[4]Other smaller datasets have been used to study mobility of specific groups. These include, among others, the Federal Reserve Bank of New York Consumer Credit Panel (DeWaard, Johnson, and Whitaker 2019), the National Longitudinal Survey of Youth, and the Survey of Income and Program Participation SIPP (Johnson and Schulhofer-Wohl 2019).

Figure 2
Inter-Commuting Zone Annual Mobility, 1980–2017



Note: The graph represents the population moving across commuting zone borders each year as a percentage of the residents in the commuting zone of destination between 1980 and 2017. The black solid line is constructed from American Community Survey data: it includes the whole US resident population (excluding those residing in group quarters) and is based on the information provided by individuals about their PUMA of residence in the previous and current year (which are matched to commuting zones of residence following the procedure of Autor and Dorn 2013). The grey solid line is constructed from the Internal Revenue Service Statistics of Income data. The series is based on the number of tax exemptions who move across counties in the previous year relative to total population of tax exemptions in the county of destination (then aggregated to commuting zone). Commuting zones are identifiable in American Community Survey data only in the years 2005–2017; the IRS series stops in 2009 because of migration misreporting acknowledged by the IRS-SOI.

market—or even the same job. In general, state-level economic statistics are likely to be an imperfect proxy for local labor markets. Figure 2 shows the total mobility rate across *commuting zones*, which have the desirable feature of proxying for self-contained local labor markets (Tolbert and Sizer 1996; Autor and Dorn 2013). As a result, migration rates between commuting zones should better capture labor demand-driven mobility as well as mobility in response to economic shocks, rather than changes of residence due to change in family status or transitions in family life (for critical discussions of the definition of local labor markets, see Monte, Redding, and Rossi-Hansberg 2018; Manning and Petrongolo 2017). In Figure 2, this mobility rate is calculated, mirroring the definition for state-mobility as those people moving into commuting zones in a given year as a percentage of the non-migrating population in the destination commuting zone.

However, measuring migration between commuting zones presents several challenges, as they are not directly observed in the Census Bureau data or in the

IRS data. For the American Community Survey line shown in Figure 2, we follow Autor and Dorn (2013) and start with the Public Use Microdata Areas (PUMAs), the smallest geographical unit available in the American Community Survey. These are geographically contiguous groups of at least 100,000 people, built by combining Census tracts and counties. The PUMAs data is available from the American Community Survey starting in 2005. We build a probabilistic match between PUMAs and the commuting zone to which they belong. We can then measure movements between commuting zones at a one-year interval.

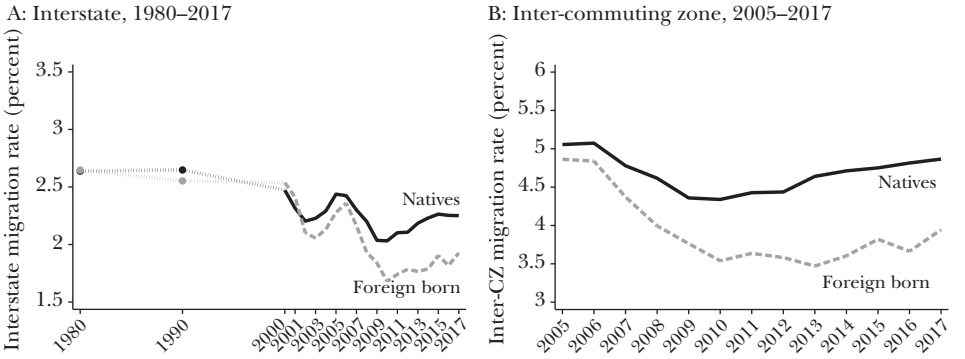
In the IRS data, locations are available at the county level, so we instead aggregate county-to-county flows into commuting zone flows. To do this, we aggregate total county in-migration flows at the commuting zone level and we subtract those coming from other counties within the same commuting zone. The IRS does not report flows below 10 units, so we need to make an assumption on those county-to-county flows that are undisclosed: in particular, we attribute them to be flows from outside the commuting zone of interest.[5]

Given that these calculations have the possibility of introducing measurement error, it is reassuring to see that the estimated migration rates in Figure 2 have similar behavior over time as those reported in Figure 1. For example, the annual rate of total inter-commuting zone migration as measured by IRS data was about 4.5 percent in early 1980s, slowly declining to around 3.5 percent in 2009. The IRS and American Community Survey series trend similarly between 2005 and 2009. Migration rate had a local peak around 2005 and then declined—by 0.8 percentage points—to the lowest level in the twelve-year period of the American Community Survey data in 2010. This is a significant decline, equal to a reduction by about one-sixth of the 2005 mobility. In the 2005–2010 period, the cross-state mobility declined by 0.4 percentage points, which also represents between one-fifth and one-sixth of the 2005 mobility. Mobility has recovered since then, although it is still below the 2005 level. Overall, while the decline in labor-market and cross-state mobility was substantial during the Great Recession, the following recovery of mobility puts the recent values in line with a continued long and slow decline in internal mobility since 1980 that affected equally long- and short-range movements.[6]

[5]Additional details are available in the accompanying Data Appendix that includes replication data and programs.

[6]Another way to measure total internal mobility is to look at lifetime interstate migration for population in working age. This can be observed in Census data by comparing state of residence and state of birth for individual in working age (that is, 15–64 years old). The Census has gathered the data necessary to calculate this mobility rate since before 1900. However, this measure is rather coarse: it does not account for when the migration occurred or for how many times a person moved in a lifetime, and it treats migration and return as no migration at all. However, this measure is qualitatively similar to the other findings: that is, the share of people who moved across states during their lifetimes increased at a slow pace until 1990, stabilized, and has slowly declined since 2000. In the most recent period, lifetime migration went from 32.0 percent in 2005 to 31.6 in 2017. For details and a figure, see the online Appendix (Figure A1) available with this article at the *Journal of Economic Perspectives* website.

Figure 3
Interstate and Inter-Commuting Zone Annual Mobility, US Natives and Foreign-Born Individuals



Note: Panel A shows the immigration rates across state borders as a percent of the state resident population: black dots for decennial Census 1980 and 1990, and black solid line for American Community Survey 2000–2017, indicates for US natives; grey dots for decennial Census 1980 and 1990, and grey dashed line for American Community Survey 2000–2017, indicates foreign-born residents of the United States. Panel B shows the immigration rates across commuting zone borders as a percent of the commuting zone resident population, separating US natives (black solid line) and foreign-born residents (grey dashed line).

Demographic and Foreign-Born Patterns of Mobility

The data from the American Community Survey allow us to focus on mobility patterns of different demographic groups. We first use it to show the difference in mobility between natives and foreign-born individuals and its evolution over time. Then we decompose mobility across groups to see the extent to which the different demographic composition of natives and immigrants can explain differences in their internal mobility, and its changes over time.

Panels A and B of Figure 3 were created using the same methods as Figures 1 and 2 and show that interstate and inter-commuting-zone mobility of natives and immigrants was similar in the period 1980–2000. After that, total foreign-born mobility seems to have declined relative to that of natives, starting in the period 2006–2010, roughly coinciding with the Great Recession. Part of this trend could be due to the drop of foreign-born arrivals from abroad, which declined from 3.5 percent of the foreign-born population in the period 2000–2007 to 2.8 percent in the period 2008–2017 (see last row of Table 1). Moreover, by comparing Figures 1 and 2 with Figure 3 panels A and B, we can see that foreign-born individuals did not affect total mobility much up to 2005, as the mobility rate of natives was very similar to overall mobility. After 2010, immigrants slightly contributed to reduce overall mobility, as in 2017, cross-state mobility of natives was about –0.1 percentage points lower than overall mobility including immigrants.

Table 1

Annual Total Migration Rates across States by Demographic Group: Natives and Foreign Born *(percent)*

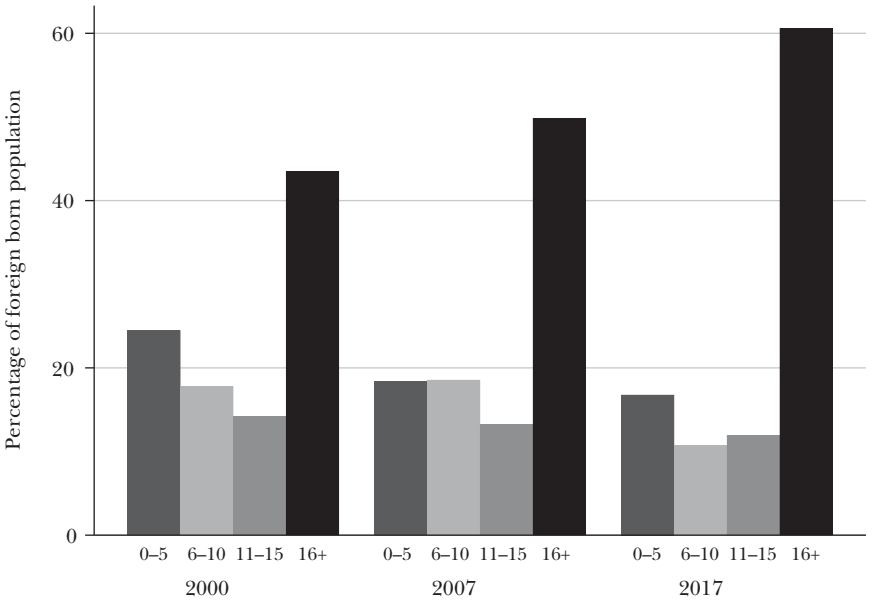
		2000–2007		*2008–2017*	
Group		*Natives*	*Foreign born*	*Natives*	*Foreign born*
Gender	*Male*	2.43	2.49	2.18	2.02
	Female	2.30	2.20	2.10	1.82
Age	*<25*	2.72	3.06	2.38	2.75
	25-44	3.22	2.94	3.16	2.6
	45-64	1.49	1.40	1.40	1.11
	65+	1.12	1.03	1.14	0.94
Race	*Other*	2.33	2.60	2.04	2.17
	White	2.37	2.09	2.17	1.66
Education	*HS*	1.87	1.85	1.66	1.33
	College	3.00	3.07	2.66	2.62
Years since arrival	*0–5*		3.95		4.03
	6–10		2.55		2.38
	11–15		2.07		1.75
	16+		1.71		1.34
Overall		2.36	2.35	2.14	1.92
One-year immigration from abroad		0.20	3.48	0.23	2.76

Source: Author's calculations based on American Community Survey data
Note: Annual total migration rates across states are calculated as the percentage ratio between in-migrants and the resident population in a given year: that is, we focus on people who change their residence between two consecutive years. The sample is composed of the US population excluding group quarter residents. Annual interstate mobility is the percent of people in each group crossing state borders in the previous year as a percent of the population of that group resident in the destination state in the previous year. In the last row, one-year immigration from abroad indicates immigration by natives and foreign-born individuals who were residing abroad in the previous year to the United States.

To see whether different demographic characteristics of immigrants and natives may explain their different recent mobility rates, Table 1 shows the differences in interstate mobility of foreign-born individuals and natives separately for several demographic groups. We report average annual mobility in the 2000–2007 period, before the Great Recession and in the 2007–2017 period that includes the Great Recession and the recovery.

Several patterns emerge from this table. First, mobility declined between the earlier and later period by about 0.2 percentage points for natives, and by 0.4 percentage points for foreign-born individuals. These changes were widespread by age, race, and education, for both immigrants and natives. Second, mobility of natives was slightly higher than foreign-born mobility on average in most groups, and this difference increased somewhat in the later period as foreign-born individuals became less mobile. Third, there are few exceptions: individuals below age 25 and who are low educated (high school and less) are more mobile for the foreign-born group than for natives.

Figure 4
Foreign-Born Residents, Distribution by Years since First Arrival in the United States



Note: Authors' calculations on 2000, 2007, and 2017 American Community Survey data, based on the reported years of residence in the United States (foreign born only). The four categories indicate respectively: 0 to 5, 6 to 10, 11 to 15, and 16 or more years of residence.

In addition, interstate mobility among foreign-born individuals is highly differentiated according to the number of years the individuals have resided in the United States, as shown near the bottom of Table 1. For example, the group that arrived in the previous five years has an incredibly high cross-state mobility rate (about 4 percent) in both the earlier and later period. This implies that newly arrived immigrants are more likely to follow opportunities and relocate, thus taking advantage of the fact that they have not yet laid down roots in a place and that they have social and human capital that is not specific to their first location. Recent flows of immigrants from abroad, therefore, are likely to represent the population most responsive to economic opportunities, both in the first move when they arrive and in a possible second move within a few years.

A decline in arrivals of new immigrants may have contributed to the decline in average mobility of immigrants. Figure 4 provides evidence on this significant change in the foreign-born population from 2000–2017: the share of foreign-born individuals who arrived more recently is falling, while the share of foreign-born individuals who have been in the United States for a longer period has risen.

Our calculations from the American Community Survey suggest that 16.8 percent of the foreign-born population in 2017 had arrived in the previous five years, while 24.5 percent of foreign-born individuals had arrived in the five years previous to 2000. Such a significant change in composition, which may also be correlated with the aging of immigrants, contributes significantly to the declining internal migration rate seen among the foreign-born.

To investigate the role of demographic compositional change among natives and foreign-born individuals, we perform a Blinder-Oaxaca decomposition to explain the decline of total mobility between 2000 and 2017.[7] This exercise decomposes the change in interstate mobility of natives and foreign-born individuals into a part due to the change of mobility for each demographic group and a part attributable to the changing shares of demographic groups in the population. We focus on individuals aged 25 to 64 who are not in school. For the period 2000–2007, demographic composition explains little to none of the change in migration patterns for both immigrants and natives. For the 2007–2017 period, changes in demographic composition taken alone would actually predict an increase (rather than the observed decrease) in migration rates for both natives and foreign-born individuals, and thus makes the observed decrease in mobility stand out even more.

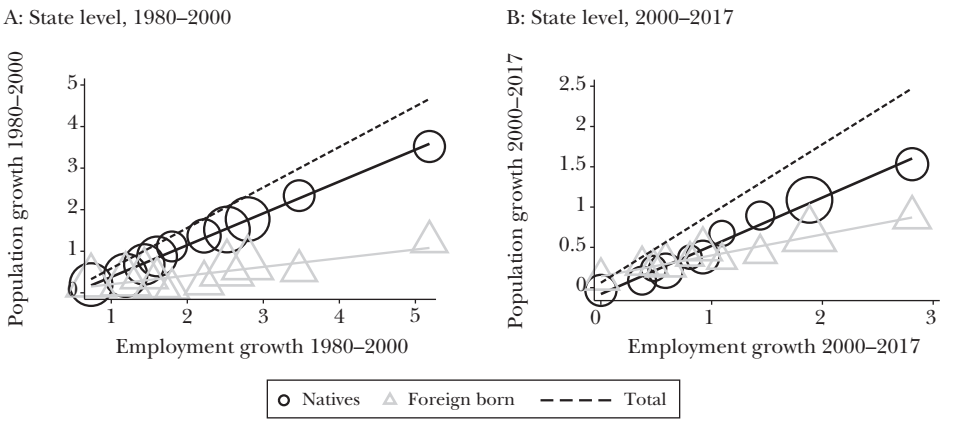
Focusing on foreign-born individuals who experienced the relatively larger decline in mobility, a decomposition based only on length of stay in the United States shows that about one-quarter of the mobility declines in the period 2000–2007 and one-half in the 2007–2017 period is due to the declining share of immigrants recently arrived (that is, in the last ten years), along with the increasing share of individuals who have been in the United States for more than 16 years.

The Responsiveness of Mobility to Local Employment Growth

Except for higher total mobility among recently arrived foreign-born individuals, total mobility seems relatively similar between foreign-born individuals and natives. Both groups show a continuing trend of declining total mobility. However, the picture changes when we look at moves in which people leave places where labor demand (and hence employment) is declining, and conversely, move toward places where labor demand is growing. That foreign-born people contributed very significantly to the population response to local labor demand shocks, improving the functioning of local markets, was previously observed by Borjas (2001). Recently, this role of foreign-born individuals in local labor markets has gained renewed attention in relation to both short-run and long-run adjustments: for US evidence, see Cadena and Kovak (2016) and Amior (2019); for European evidence, see Basso, D'Amuri, and Peri (2019).

[7] Detailed results of the decomposition are in the online Appendix available with this paper at the *Journal of Economic Perspectives* website.

Figure 5
Population Change and Employment Growth, Yearly Average *(percent)*



Note: The graphs show the average annual change in the native population and foreign-born population, as percentage of the total initial population, in response to the average annual percentage change of total employment in US states. The scatter represents the average mobility in each of ten deciles of employment growth and the size of the bubble is proportional to the sum of the state population in each decile at the beginning of the period. The regression lines represent the linear relationship between the change in the total (dark-dashed), native (dark solid) and foreign-born (light solid) population as percent of initial total population, and the change in total state employment in each period. The sample is composed of 25–64 year-olds not enrolled in school and not residing in group quarters. Panel A reports these figures for the period 1980–2000 using decennial Census data and Panel B for the period 2000–2017 using the 2000 decennial Census and the 2001–2017 American Community Survey data.

Figure 5 shows the relationship between population change (vertical axis) and employment growth (horizontal axis) based on state-level U.S. data. A coefficient of one—shown by the dashed line in the two panels—would imply that employment changes, likely driven by local labor demand changes, were fully accompanied by changes in the specific population of similar magnitudes, leaving the employment-population ratio unchanged. A coefficient of zero, on the other hand, would imply no net population changes and only changes in local employment-population ratios in response to employment shocks. Analyzing the relationship between total population change and total employment change illustrates the extent to which employment changes were associated in aggregate with net inflows or outflows of people. Using this approach, and separating the US resident population into foreign-born and natives, we can see which part of the total population adjustment is driven by changes in natives and which part by changes in the foreign-born population.

One way to represent local population adjustment due to a specific group, associated with changes in local labor demand, is to plot changes in that population as a percent of the initial total population versus percentage changes in employment in the same geographic units. While for simplicity we only show a correlation, empirical evidence by Cadena and Kovak (2016) during the Great Recession and

by Amior (2019) for the long-run—which include controls and exploit variation of sector-specific shocks to better isolate demand factors—find estimates consistent with our basic results.

More specifically, Figure 5 divides the states into deciles based on the growth rate of employment in the period 1980–2000 in Panel A, and in the period 2000–2017 in Panel B. In each panel, we show the correlation line for the percentage change in total population versus percentage employment growth as a dark dashed line. We also show the change in native population as a percentage of the total, represented by the round circles and continuous black line, and changes in the foreign-born population as a percentage of the total, represented by the triangles and the grey line. Each circle or triangle represents a group of five states, binned into the same decile of the employment growth distribution. The slope of the line represents the association of population growth with employment growth, and the sum of the slopes for the native and foreign-born populations equals the slope of the total population response.

Considering the changes across states, represented in Panel A of Figure 5, we see that the overall population responded essentially one-for-one to employment during the 1980–2000 period: the estimate of that slope is equal to 1.01.[8] Analyzing the contribution of each group, 73 percent of population adjustment in response to employment changes was due to natives, and 27 percent was due to foreign-born individuals. Only about 11 percent of the population was foreign-born during this period, and the foreign-born population accounted for 10.8 percent of total interstate annual mobility. Thus, about one-fourth of state-level population adjustment in response to employment changes was due to foreign-born internal migration in the period 1980–2000 when the foreign-born represented only one-tenth of the US population and of total interstate mobility. This implies that the foreign-born population was about 2.5 times more responsive than the native population in moving to locations experiencing positive economic shocks and away from those experiencing negative economic shocks.

Evidence from the period 2000–2017 confirms the main features described for 1980–2000 but with an additional twist. In Panel B of Figure 5, the overall population response to employment changes decreased somewhat in this period relative to 1980–2000, so that only 86 percent (rather than 101 percent, as in 1980–2000) of the employment change was adjusted by population changes. During this period, the share of foreign-born population increased to about 17 percent and the foreign-born share of total interstate mobility was around 16 percent. The foreign-born population contribution to the local employment growth response represented 36 percent of the total population adjustment—again, more than twice

[8] Table A2 in the online Appendix, available with this article at the *Journal of Economic Perspectives* website, shows the estimates of the slopes, capturing the population change associated to employment changes, and additional details of data and regression results presented throughout this discussion of Figure 5. Notice also that the regression includes an intercept allowing the average national population growth rate and the average national employment growth rate to be different. Both in 1980–2000 and in 2000–2017 employment of 25–64 years old grew faster than population and the national employment/population ratio increased nationally. This does not affect the cross-state estimates of these slopes.

the population share of the foreign-born population. In the 2000–2017 period, which included a deep recession and recovery, the mobility of natives associated with employment shocks decreased, while the population response of immigrants remained as strong as in the previous decades. This implies that as the foreign-born population share also grew, they became responsible for more than one-third of the local population adjustment across US states associated with employment changes.

The correlations shown in Figure 5 continue to hold if we add various controls: the past economic conditions of a location, such as the employment/population ratio at the beginning of the period (as in Amior 2019) and if we consider yearly or long-run changes over decades. Basso, Peri, and Rahman (2017) analyze the difference in mobility between native and foreign-born individuals in response to a more direct measure of labor productivity shocks (related to the adoption of computers). They also find a larger response of foreign-born population relative to native population, when looking at the period 1980–2010, especially among less educated individuals. Their evidence implies that the share of population adjustment of foreign-born individuals in response to local employment changes was two times larger than their share of the overall population.

This greater propensity of foreign-born individuals to move toward areas with increasing employment growth is particularly interesting if we add two qualifications that have been emphasized in the literature. First, the largest difference in mobility is between less-educated (say, non-college educated) native and foreign-born individuals (Cadena and Kovak 2016). While college-educated natives and foreign-born individuals have greater and similar mobility in response to employment growth, the population of non-college educated natives has become much less responsive over time to employment growth, while foreign-born, non-college educated individuals were and remained quite responsive to it.

Second, the tendency of less-educated natives to become less mobile in response to employment and wage opportunities across locations is a trend that has continued for the last several decades. Ganong and Shoag (2017) show that net mobility of less-educated individuals into high-income locations in 1935–1940 was positive and much greater than in 1995–2000, when the less-educated (differently from the more-educated) did not exhibit any tendency to move towards high-income locations (see their Figures 4 and 5). They attribute a large part of the decline among the less-educated in their tendency to move toward employment and wage opportunities to the change in local prices (especially for housing) that in many successful locations increased substantially. In their argument, the higher local prices more than offset the income gains available to the less-educated, but not for the more highly educated, whose share of income spent on housing was smaller. Moreover, they connect the disproportionate increase in housing prices in many successful and desirable locations with restrictions to housing supply caused by regulation and natural constraints and argue that such regulations are a large hurdle to mobility, especially for the less-educated, and hence are a hurdle to convergence of wages/income across locations.

An important corollary to the decline in the mobility of the less-educated toward wages and jobs relative to the highly educated is that economically successful

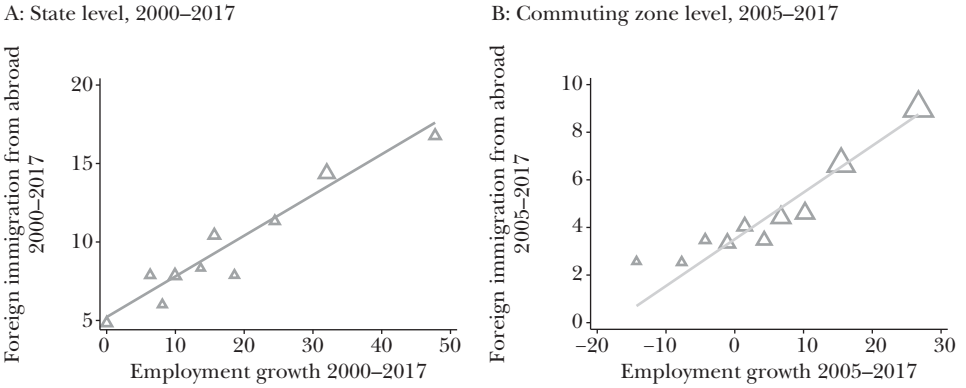
cities and regions—those where employment opportunities and wages grew—have become places with higher concentrations (or shares) of highly educated individuals. This fact has deep economic consequences, which have been recognized and studied by several urban economists and particularly emphasized by Moretti (2012, 2013). Economically successful cities attract more highly educated individuals and this, in turn, feeds their economic success, creating further innovation and productivity growth. At the same time, this dynamic increases housing prices in those locations, especially in many highly successful cities such as New York, Seattle, and San Francisco where housing regulations do not allow an expansion of the housing stock. This high price of housing contributes to keeping out the less-educated. At the same time, higher concentrations of college-educated people may contribute to increasing local amenities (such as quality of schooling, art, or variety of consumption) that are especially valued by college-educated individuals relative to non-college educated individuals (Diamond 2016). This further enhances the cycle, attracting the highly educated more than the less-educated.

There is an interesting implication of the facts presented above. Highly populated and successful cities have increased their density (share) of both foreign-born individuals and highly educated natives (Peri 2016). Albert and Monras (2018) show that the elasticity of immigrant share to population density across cities is positive, very significant, and increased during the period 1980–2010 when, as we show above, the foreign-born population responded significantly more to employment growth. Highly successful cities where wages and population density, local prices, and housing values are high, were places where, especially in the 1980–2000 period (but also in the 2000–2017 period), employment grew faster and the share of foreign-born individuals increased, including those with low levels of education.

Why were foreign-born individuals much more willing than natives to move to fast-growing, higher-wage, but increasingly also more expensive labor markets during the last four decades? Before addressing this question, let us note one additional important fact. A significant part of the mobility of foreign-born individuals to fast-growing labor markets comes from recently arrived immigrants (Amior 2019; Cadena and Kovak 2016). In particular, the first location of immigrants upon arrival from abroad as well as their early relocations within the first five years since migration (when foreign-born individuals exhibit an extremely high *total* mobility as shown in Table 1) account for the largest part of foreign-born mobility to locations with higher employment growth.

Figure 6 shows the correlation between foreign-born cumulated new immigrant arrivals from abroad as a share of the initial population and local employment growth, with Panel A showing mobility at the state level from 2000–2017 and Panel B showing mobility at the commuting-zone level for 2005–2017. Again, employment growth is on the horizontal axis. However, the vertical axes now measures the immigration rate only of newly arrived foreign-born individuals—that is, those coming directly from a residence abroad who were not residing before in the United States over the full period (not the population growth of the foreign-born as in the earlier Figure 5). Again, the triangles represent five states each, grouped by decile of employment

Figure 6
New Immigrant Arrival Rate and Employment Growth, Cumulative *(percent)*



Note: The graphs show the cumulated 1-year new immigration rate (in percent of the base year population) of foreign-born population at the state-level over the period 2000–2017 (Panel A) and at the commuting zone level over the period 2005–2017 (Panel B). On the *x* axis there is the area employment growth over the relative period in percent. The scatter represents the average immigration in each of ten deciles of employment growth and the size of the triangle is proportional to the sum of the state population in each decile at the beginning of the period. The regression lines represent the linear relationship between the cumulative new immigration rate and the change in total state employment in each period. The sample is composed of 25–64 year-olds not enrolled in school and not residing in group quarters and is drawn from the 2000 decennial Census and the 2001–2017 American Community Survey data.

growth, with the size of the triangle proportional to the population of that group. We see a positive and very significant correlation that implies a net increase in the inflow of newly arrived immigrants of .26 percentage points for each 1 percent increase in state employment growth. The mean increase in state-level employment during this time period (about 17 percent) would move a state from the median to the seventy-fifth percentile of the immigration rate of newly arrived immigrants (coming directly from abroad). Similarly, a 1 percent growth in employment at the commuting zone level was associated with a .25 percentage point increase in immigration rate of newly arrived immigrants. In spite of relatively small inflows of new immigrants from abroad in the more recent period considered—and especially in the period 2006–2010 during the Great Recession (Cadena and Kovak 2016)—new foreign-born migration was quite significant in response to local employment growth.

What Explains the High Responsiveness of Foreign-Born Mobility to Economic Conditions?

What can explain the much larger migration of the foreign-born toward fast-growing but dense and expensive local labor markets, vis-à-vis natives? Can we associate it with some features and choices that systematically differentiate

foreign-born individuals from natives? We provide five different potential explanations listed according to our assessment of their importance.

The first explanation is strictly connected to the role played by local prices (especially of housing) and their steep increases in reducing the real gains of less-educated workers in booming cities (Ganong and Shoag 2017). This explanation, proposed by Albert and Monras (2018), focuses on the idea that because foreign-born people plan to spend significant shares of their permanent income in their countries of origin rather than where they live—in the form of remittances, transfers to families, and future consumption there if or when they return—they face an effective "price index" which is less sensitive to local prices. In support of this theory, Albert and Monras show that foreign-born people from countries with lower price indices are more likely to live in high-wage, high-price locations in the United States. They also show that more recent immigrants who have higher probabilities of return and higher shares of income remitted to their countries of origin are also more likely to live in high-wage, high-cost cities.

We would add a variation to this explanation: foreign-born individuals may be less sensitive to local housing prices because they consume less housing in terms of space and quality. The foreign-born may be more used to dense living arrangements in their countries of origin or more used sharing a housing unit with extended families, if needed. Padovani (2018) shows that foreign-born (but not native) households have higher densities of persons per room in dwellings in high-price cities relative to low-price cities. Another explanation based on housing markets may help rationalize the higher propensity of recent migrants to leave locations that show economic decline. A decline in the local housing market, generating negative equity value in a house, can make homeowners less likely to move (for example, Ferreira, Gyourko, and Tracy 2010). However, immigrants, especially those recently arrived, are much less likely to be homeowners and hence are much less likely to be caught in a negative-equity trap. The housing equity channel may explain some decline in mobility in the Great Recession and some differences between natives and immigrants during that period. However, the explanation does not seem important in explaining the long-term decline in native mobility (Winkler 2011; Molloy, Smith, and Wozniak 2011) and may only be marginally important in accounting for differences in the differential mobility of foreign-born individuals.

The second explanation focuses on the well-known tendency of new immigrants to locate where previously established communities of immigrants from the same country are already established. This "enclave" theory (Card 2001) suggests that the location of foreign-born individuals by nationality is persistent over time. In order for this tendency to produce migration of newly arrived immigrants toward more economically successful cities during the 1990s and 2000s, it must be the case that early immigrants—say, those who arrived in the United States in the 1970s and 1980s—had previously located in these cities. Amior (2019) argues for this explanation. He emphasizes that demand shocks for labor are persistent over time and hence a strong response of immigrants to local economic success in the early years is positively correlated with employment shocks in the following decades.

This explanation seems consistent with some basic correlations in our sample for the period 2000–2017, but much less so in the 1980–2000 period. In particular, during the 1980–2000 period, the excess population adjustment of foreign-born persons (shown in the previous section) derived from a much higher elasticity of own-population response of the foreign-born population to local employment changes relative to natives. Essentially, in these early decades of large international migration, new immigrants were very responsive to local employment growth. We estimate an elasticity of own population to employment of about 4 for the foreign-born vis-à-vis an elasticity of only 0.8 for natives. Before this period, the shares of immigrants in US labor markets were small, as immigration was low in the 1960s and 1970s. Hence, as demand shocks affected cities and states differentially, the newly arrived migrants of the 1980s and 1990s were very responsive to them and created new "enclaves" of immigrants across the United States.

Then, in the period 2000–2017, when some cities had reached high densities of immigrants while other had not, the disproportionate role of foreign-born population growth in response to employment growth was actually due to a significant positive correlation between the share of immigrants in 2000 and the subsequent employment growth. While during this period, the own-population elasticity of the foreign-born to employment became similar to that of natives and rather small, the constituted immigrants' enclaves both attracted new immigrants and were also the locations with more significant economic success and employment growth in the 2000–2017 period. Whether the correlation across labor markets of the foreign-born share in 2000 and employment growth in 2000–2017 was only due to the persistence of employment shocks, or whether immigrants contributed to it, by promoting productivity gains or innovation as argued in Kerr and Lincoln (2010) and Peri (2012), it should be a subject of further investigation. Here, we only capture a correlation between the earlier location of immigrants and later employment growth across states. Each one percentage point greater share of the immigrant population in 2000 was associated with 0.7 percent faster employment growth over the 2000–2017 period.

The third explanation is based on a key difference in the type of jobs (occupations) taken by foreign-born individuals and natives. In particular, foreign-born individuals—especially those who have recently arrived—are much less likely to be in occupations where licensing is important, such as real estate brokers, pharmacists, physical therapists, physicians, or electricians. Cassidy and Dacass (2019) show that immigrants are much less likely to be licensed than natives. Johnson and Kleiner (2017) show that occupations with state-specific licensing inhibit interstate mobility significantly. Using data from 2005 to 2015, they show that people in occupations with state-specific licensing had a probability of interstate mobility 32 percent lower than people in occupations with no licensing requirements. Hence, the state-specificity of licensing may contribute to reduce the responsiveness of natives, who are more likely to be employed in licensed occupations, to out-of-state employment growth. Foreign-born individuals, however, are less affected as they are less likely to be in

such occupations. Occupations with very high share of licensed workers such as firefighters and paramedics (67 percent and 80 percent of which are licensed) tend to have a low share of the foreign-born among them (3.5 and 5.3 percent respectively). To the contrary, occupations such as drywall installers and housekeepers, with only 3 percent of the group licensed, include 60 and 52 percent of immigrants among them, respectively.[9] This explanation, however, does not seem specific to a greater proportion of foreign-born moves directed to high-income/high-price cities, nor should it affect much mobility within a state. Moreover, licensing should not affect less-educated natives more than highly educated ones, as being employed in occupations which heavily rely on licensing is more common among college-educated than among non-college-educated individuals.

A fourth explanation, not yet carefully explored, is also related to the different occupations of natives and foreign-born individuals, but along a different dimension. Recent technological evolutions have increased the creation of high-skilled, cognitive-intensive jobs, but also of manual, non-routine jobs, while reducing the creation of routine-intensive types of jobs (Autor 2019). Those manual non-routine jobs (in personal service, construction, house services) have been taken in large part by immigrants (Basso, Peri, and Rahman 2017; Mandelman and Zlate 2016). Newly arrived, low-skilled foreign-born individuals may have comparative advantages in those manual jobs for several reasons: their language proficiency is limited (Peri and Sparber 2009) and they may have more tolerance and less distaste for such jobs, as a consequence of the higher status of these jobs in their countries of origin. If those manual and non-routine jobs were disproportionally created in dense, high-income areas, where technology and high-income consumers generate demand for them, this could explain the higher demand for recently arrived foreign-born people in those areas.

Finally, internal mobility in the United States may have decreased because of a decline in geographic specificity of earnings for skills (or occupations) and because of an improvement in workers' ability to learn about economic returns in a location before moving, hence reducing multiple migration (Kaplan and Schulhofer-Wohl 2017). One might argue that because of higher uncertainty in evaluating the skills of the foreign-born, and because the foreign-born have less information about local opportunities, multiple migration might be expected to be greater for this group. Such an explanation, however, seems more suitable to explain a difference in overall mobility, as some lack of information may generate random mobility rather than mobility in response to economic differences across labor markets. It can, however, help us understand the initially high total mobility

[9] In the online Appendix, Figure A2 shows the correlation between the share of people licensed and the share of foreign-born people across state-occupation cells in year 2016. The scatterplot shows the share of foreign-born persons in state-occupation cells on the vertical axis against the share of the licensed workforce on the horizontal axis, ranked from lowest to highest licensing share and binned into cells each including 5 percent of observations. We see a strong negative correlation with a much higher presence of foreign-born people in cells with less than 20 percent of licensed workers. We thank Joshua Grelewicz for making his data on licensing and foreign-born persons available to us and for assisting us in using them.

of recently arrived immigrants relative to long-term residents, as uncertainty and a lack of information is more severe in the initial period after arrival to the United States. Recently arrived immigrants can also be more mobile as they have less social connections and less location-specific social and human capital, which makes it less costly for them to move again.

Consequences and Policy Discussion

Whatever the reasons for the high mobility of foreign-born individuals in response to local economic shocks, we conclude this essay by looking at the implications for the native-born, and particularly what policy considerations arise in light of these consequences.

First, high mobility of the foreign-born in response to local economic shocks is beneficial to the native-born as a group. Whether it ensures greater local employment adjustment (as argued in Cadena and Kovak 2016) and therefore reduces wage fluctuations for natives in response to shocks, or whether it simply replaces native mobility, allowing natives to avoid moving and displacement costs during economic downturns (as argued by Amior 2019), foreign-born individuals constitute a "buffer" that reduces adjustment costs to labor demand shocks for natives and ease the burden of short-run adjustment for natives.

Second, the higher propensity of foreign-born individuals to move to more productive, more expensive cities increases the overall efficiency of labor allocation in the United States (as also shown by Albert and Monras 2018). In short, it moves productive resources of people and their labor to cities where productivity is higher. On the other hand, when foreign-born individuals help to provide manual services that high-income cities need, it may also provide those cities with a means to lower the cost of non-tradable services without removing housing regulations. In turn, however, this may contribute to the crowding out of some less-educated natives from those cities.

In light of these consequences for natives, we think three immigration policies may potentially enhance the positive effects of immigrants on native welfare. First, because the flow of new immigrants is responsive to employment changes in US labor markets, we could consider making immigration quotas more responsive to US employment cycles. This would increase the aggregate US immigration in periods of high employment growth and reduce it at other times, working as an aggregate adjustment mechanism. Second, as the foreign-born concentrate in high-productivity/high-price cities, increasing their income but also contributing to the crowding of some local services (like schools), it may be prudent to transfer some of the federal income tax gains from foreign-born workers back to those cities to be invested in local services.

Finally, because large, dense, and highly productive cities attract a much larger share of foreign-born individuals than do other places and are thus more affected by immigration than other cities, it seems sensible that the governments of those

cities (perhaps via their mayors) should have more prominent input into federal immigration policies. For example, these mayors may have useful input on issues like the prioritization of enforcement, incentives, and requirements for entry and rules on the number and types of immigrants. As immigrants—especially newly arrived ones—play a very important role in how local labor markets respond to labor demand shocks, local and regional considerations should play a more explicit role in national immigration policies.

■ *We thank Gordon Hanson, Enrico Moretti, Heidi Williams, and Timothy Taylor for very useful comments and suggestions on a previous draft of the paper. The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Bank of Italy nor of the Eurosystem. Any errors or omissions are the sole responsibility of the authors.*

References

Albert, Christoph, and Joan Monras. 2018. "Immigration and Spatial Equilibrium: The Role of Expenditures in the Country of Origin." CEPR Discussion Paper 12842.

Amior, Michael. 2019. "The Contribution of Foreign Migration to Local Labor Market Adjustment." CEP Discussion Papers 1582.

Autor, David H. 2019. "Work of the Past, Work of the Future." *American Economic Association: Papers and Proceeding* 109: 1–32.

Autor, David H., and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103 (5): 1553–97.

Basso, Gaetano, Francesco D'Amuri, and Giovanni Peri. 2019. "Immigrants, Labor Market Dynamics, and Adjustment to Shocks in the Euro Area." *IMF Economic Review* 67 (3): 528–72.

Basso, Gaetano, Giovanni Peri, and Ahmed Rahman. 2017. "Computerization and Immigration: Theory and Evidence from the United States." NBER Working Paper 23935.

Borjas, George J. 2001. "Does Immigration Grease the Wheels of the Labor Market?" *Brookings Papers on Economic Activity* 32 (1): 69–134.

Cadena, Brian C., and Brian K. Kovak. 2016. "Immigrants Equilibrate Local Labor Markets: Evidence from the Great Recession." *American Economic Journal: Applied Economics* 8 (1): 257–90.

Card, David. 2001. "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration." *Journal of Labor Economics* 19 (1): 22–64.

Cassidy, High, and Tennecia Dacass. 2019. "Occupational Licensing and Immigrants." Center for Growth and Opportunity Working Paper 2019-9.

DeWaard, Jack, Janna Johnson, and Stephan D. Whitaker. 2019. "Internal Migration in the United States: A Comprehensive Comparative Assessment of the Consumer Credit Panel." *Demographic Research* 41: 953–1006.

Diamond, Rebecca. 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000." *American Economic Review* 106 (3): 479–524.

Ferreira, Fernando, Joseph Gyourko, and Joseph Tracy. 2010. "Housing Busts and Household Mobility." *Journal of Urban Economics* 68 (1): 34–45.

Ganong, Peter, and Daniel Shoag. 2017. "Why Has Regional Income Convergence in the U.S. Declined?" *Journal of Urban Economics* 102: 76–90.

Grelewicz, Joshua. 2019. "Data and Elaborations on Licensing and Mobility." Unpublished Data. (accessed October 2019).

Ihrke, David. 2014. *Reason for Moving: 2012 to 2013*. Washington, DC: U.S. Census Bureau https://www.census.gov/prod/2014pubs/p20-574.pdf.

Internal Revenue Service (IRS). 1983–1989. "County to County, State to State and County Income Migration Flow Data Files." https://catalog.archives.gov/id/646447.

Internal Revenue Service (IRS). 1990–2009. "Statistics of Income Tax Stats—Migration Data." IRS Tax Statistics. https://www.irs.gov/statistics/soi-tax-stats-migration-data. (accessed October 15, 2019).

Johnson, Janna E., and Morris M. Kleiner. 2017. "Is Occupational Licensing a Barrier to Interstate Migration?" NBER Working Paper 24107.

Johnson, Janna E., and Sam Schulhofer-Wohl. 2019. "Changing Patterns of Geographic Mobility and the Labor Market for Young Adults." *Journal of Labor Economics* 37 (S1): S199–S241.

Kaplan, Greg, and Sam Schulhofer-Wohl. 2012. "Interstate Migration Has Fallen Less Than You Think: Consequences of Hot Deck Imputation in the Current Population Survey." *Demography* 49 (3): 1061–74.

Kaplan, Greg, and Sam Schulhofer-Wohl. 2017. "Understanding the Long-Run Decline in Interstate Migration." *International Economic Review* 58 (1): 57–94.

Kerr, William R., and William F. Lincoln. 2010. "The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention." *Journal of Labor Economics* 28 (3): 473–508.

Mandelman, Federico S., and Andrei Zlate. 2016. "Offshoring, Low-Skilled Immigration, and Labor Market Polarization." Federal Reserve Bank of Boston Working Papers RPA16-03.

Manning, Alan, and Barbara Petrongolo. 2017. "How Local Are Labor Markets? Evidence from a Spatial Job Search Model." *American Economic Review* 107 (10): 2877–907.

Meyer, Bruce D., and Nikolas Mittag. 2019. "Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness, and Holes in the Safety Net." *American Economic Journal: Applied Economics* 11 (2): 176–204.

Meyer, Bruce D., Wallace K. C. Mok, and James X. Sullivan. 2015. "Household Surveys in Crisis." *Journal of Economic Perspectives* 29 (4): 199–226.

Meyer, Bruce D., Derek Wu, Victoria R. Mooers, and Carla Medalia. 2019. "The Use and Misuse of Income Data and Extreme Poverty in the United States." NBER Working Paper 25907.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the US: Updated Facts and Recent Trends." *Journal of Economic Perspectives* 25 (3): 173–96.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg. 2018. "Commuting, Migration, and Local Employment Elasticities." *American Economic Review* 108 (12): 3855–90.

Moretti, Enrico. 2012. *The New Geography of Jobs*. New York: Houghton Mifflin Harcourt.

Moretti, Enrico. 2013. "Real Wage Inequality." *American Economic Journal: Applied Economics* 5 (1): 65–103.

Padovani, Andrew. 2018. "Immigrants and the Great Divergence." PhD diss. University of California, Davis.

Peri, Giovanni. 2012. "The Effect of Immigration on Productivity: Evidence from U.S. States." *Review of Economics and Statistics* 94 (1): 348–58.

Peri, Giovanni. 2016. "Immigrants Productivity and Labor Markets" *Journal of Economic Perspectives* 30 (4) 3–30.

Peri, Giovanni, and Chad Sparber. 2009. "Task Specialization, Immigration, and Wages." *American Economic Journal: Applied Economics* 1 (3): 135–69.

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6): 1257–78.

Rosen, Sherwin. 1979. "Wages-Based Indexes of Urban Quality of Life." In *Current Issues in Urban Economics*, edited by Peter Mieszkowski and Mahlon R. Straszheim, 74–104. Balitmore: John Hopkins University Press.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2019. "IPUMS USA: Version 9.0 [dataset]." IPUMS. https://doi.org/10.18128/D010.V9.0. (accessed July 31, 2019).

Tolbert, Charles M., and Molly Sizer. 1996. "US Commuting Zones and Labor Market Areas: A 1990 Update." U.S. Department of Agriculture Economic Research Service Staff Paper 9614.

Winkler, Hernan. 2011. "The Effect of Homeownership on Geographic Mobility and Labor Market Outcomes." Society for Economic Dynamics Meeting Paper 196.

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 99–127

Using Place-Based Jobs Policies to Help Distressed Communities

Timothy J. Bartik

This article focuses on place-based jobs policies: policies that provide assistance to individual businesses to encourage job growth in a particular local labor market, such as a metropolitan area. When run by state and local governments, place-based jobs policies are commonly called economic development policies. Such business assistance includes business tax incentives, cash grants to business, and special public services to business such as customized job training, manufacturing extension services, small business development centers, business incubators, business parks, and access roads.

Place-based jobs policies currently cost around $60 billion annually (as discussed in the next section). Over the past 30 years, the typical state and local incentive offer, as a percent of the value-added of the business receiving the incentive, has tripled in size. Do current policies make sense? How should they be reformed? Can reforms be carried out by state and local governments, or is federal intervention needed? These are some of the questions I will address.

One rationale for place-based jobs policies is that local labor markets have large and persistent differences in job availability. A commonly-used definition of local labor markets is commuting zones, which are groups of counties which contain most commuting flows; commuting zones are similar to metro areas, but include all US counties. Consider the distribution across commuting zones of the "prime-age employment rate"—that is, the employment-to-population ratio for 25–54 year-olds, as provided by the 2018 American Community Survey (Ruggles et al. 2020). Commuting zones at the median of the population distribution had a prime age employment rate of 80.9 percent, compared with 75.5 percent at the 10th percentile

■ *Timothy J. Bartik is Senior Economist, W.E. Upjohn Institute for Employment Research, Kalamazoo, Michigan. His email address is bartik@upjohn.org.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.99.

and 84.5 percent for the 90th percentile—a 90/10 gap of 9.0 percentage points. Moreover, spatial differences in employment rates persist for decades (Amior and Manning 2018; Austin, Glaeser, and Summers 2018).

These low employment rates in certain local labor markets have large costs both for the individual nonemployed and for the broader society. For the individual, nonemployment not only reduces earnings but leads to poorer mental health (Diette et al. 2018). Local job losses lead to increases in alcohol and drug use and to opioid deaths (Autor, Dorn, and Hanson 2018). Social costs of local labor market distress include the following five items: First, local labor market distress leads to increases in property crime (Pierce and Schott 2017). Second, such distress leads to increases in single-mother families (Autor, Dorn, and Hanson 2018). Third, lower parental income is associated with children doing worse in educational attainment and adult income (Bastian and Michelmore 2018; Stuart 2017). Fourth, declines in local employment rates cause fiscal problems—for example, increased welfare benefits such as disability (Charles, Hurst, and Schwartz 2018). Fifth, an individual's life satisfaction is negatively affected by both their own unemployment and the area's unemployment (Helliwell and Huang 2014).

In the past, researchers such as sociologist William Julius Wilson (1996) have talked about the social problems that develop when work disappears in inner cities. Today, it is widely recognized that local joblessness and its accompanying social problems have spread to many smaller cities and rural areas, as documented by writers such as J.D. Vance in *Hillbilly Elegy* (2016).

To alleviate the costs of local labor market distress, we might want to encourage businesses to create jobs in these distressed areas. The next section describes current place-based jobs policies.[1] The bulk of the article then discusses market failures that might justify place-based jobs policies. This analysis leads to suggested reforms, which could be implemented by state and local governments but might also be promoted by the right type of federal intervention.

What are Place-Based Jobs Policies?

Local-area job growth is potentially affected by many government actions. Here, I focus on state and local government interventions aimed at increasing job growth in a state or in a local labor market, often referred to as "economic development" policies. These policies include tax breaks or cash that are awarded in a discretionary fashion to individual businesses. They also include services customized to individual businesses, like customized job training for a specific firm or advice for firms provided through manufacturing extension services and small business development centers. State and local economic development policies can also include infrastructure investments and land development practices.

[1] This article does not discuss another type of place-based policy: community development policies, which seek to improve places much smaller than a metro area, such as a census tract or other small neighborhood.

Table 1
Resources Devoted to Place-Based Jobs Policies in the United States

Current programs	
Policy/program	*Annual dollars (in billions)*
State and local business tax incentives and other cash incentives	47.1
Non-financial incentives: customized training programs	0.6
Other state economic development programs	2.8
Subtotal, state/local programs	50.6
Manufacturing extension (federal/state/fees)	0.4
Economic Development Administration (EDA)	0.3
Economic development portion of HUD's Community Development Block Grants	1.1
Small Business Administration	0.8
Other economic development programs in USDA, HUD, Commerce	2.0
Subtotal mostly federal spending	4.7
Opportunity Zones tax credits	1.6
New markets tax credit	1.5
Other tax expenditures that might promote local economic development	2.4
Subtotal, federal tax expenditures	5.4
Total of federal programs and tax expenditures	10.1
Total of all levels of government	60.7
Past programs	
Appalachian Regional Commission (peak annual spending 1966–1975)	1.7
Tennessee Valley Authority (peak annual spending 1950–1955)	1.5

Note: All figures in 2019 dollars. State/local cash incentives are 2015 data from Bartik (2017a). Customized job training spending from Hollenbeck (2013). Other state economic development programs from State Economic Development Expenditure Database (Council for Community and Economic Research 2018). To avoid double-counting, I subtract out workforce preparation, strategic business attraction, and business assistance, and subtract out the half of state economic development spending which is federally-financed (Council for Community and Economic Research 2017). This spending includes: tourism, film promotion, other special industry promotion, high-tech programs, business finance, entrepreneurial assistance, minority business development, community assistance, business recruitment, and trade promotion. Manufacturing extension from T. Bartik (2018b). EDA, HUD, and SBA from FY 2017 US federal budget. For CDBG, assume one-third of funds are for economic development. Other economic development spending from GAO (2012b). Opportunity Zones from Kimbo and Phillips (2018). New markets tax credit costs from US Department of Treasury (2016). Other tax expenditures from GAO (2012a). ARC figures from Jaworski and Kitchens (2019). TVA figures from Kline and Moretti (2013).

With this focus, Table 1 summarizes current place-based jobs policies. By far the largest place-based jobs policies are state and local financial incentives to business provided via tax breaks or cash. Such financial incentives were $47.1 billion (in 2019 dollars) in 2015, almost four-fifths of the annual total for all place-based jobs policies of $60.7 billion.[2] State and local governments also provide non-financial

[2]Some estimates of incentives are larger: Thomas (2011) calculates $73 billion and Story (2012), $101 billion (in 2019 dollars). However, Thomas includes tax breaks that are not aimed at job growth. Story includes exemptions of business inputs from sales taxes, generally regarded as a desirable tax feature. Slattery and Zidar (2020) get lower estimates at $22 billion (2019 dollars) for state incentives only, based on state budgets and tax expenditure reports.

incentives through customized job training programs under which community colleges provide free training to firms locating or expanding a facility—but these non-financial incentives are only $0.6 billion. I now focus on describing how state and local incentives are designed and distributed. This description is based on Bartik (2017a), where I use a hypothetical firm model and construct a database which calculates incentives paid over 20 years for a new facility (started up in years 1990-2015) in each of 45 industries and 33 states. The 33 states make up 92 percent of US GDP; the 45 industries make up 91 percent of US compensation.

To put the annual incentives of $47.7 billion in perspective, this total is almost identical to the $47 billion in state corporate income taxes (US Census Bureau 2020), which sounds like a lot. On the other side, $48 billion is less than 3 percent of state and local own-source tax revenue. As we will discuss, state and local governments target incentives at businesses selling outside the state—the "export-base" or "tradable goods" sector. For firms in export-base industries, typical incentives offset 30 percent of a business's state and local taxes. But the typical incentive package equals only 1.4 percent of the business's value-added, or 3.1 percent of wages for such businesses.[3]

Although most incentives are provided by state governments, 27 percent come from local governments through property tax abatements. The largest incentive is job creation tax credits, typically defined as a percentage of the increased wage bill at the new or expanded business, which make up 45 percent of total incentives. Many job creation tax credits are refundable—that is, they can exceed the firm's state corporate income tax liabilities. Some states allow a new or expanding firm to keep the workers' state income tax withholdings associated with the new jobs. More minor roles are played by investment tax credits (14 percent) and research and development (R&D) tax credits (9 percent). Customized job training incentives are less than 5 percent of total incentives.

The share of incentives received by large firms is disproportionate. Firms with over 100 employees get more than 90 percent of incentives (Chatterji 2018), while firms of this size represent only 66 percent of private jobs (based on the Longitudinal Business Database for 2016, from US Census Bureau 2018). Incentives particularly go to the very largest firms. Slattery and Zidar (2020) find that for new establishments with more than 1,000 employees, over 36 percent receive incentives, versus less than 10 percent for establishments between 500 and 999 employees, less than 2 percent for establishments between 250 and 499 employees, and even lower percentages for smaller establishments.

As mentioned, incentives are usually targeted at "export-base industries"—industries that sell their goods and services outside the state. In 2015, in the average state, the present value of incentives, as a percentage of the present value of value-added, was 1.42 percent for export-base industries versus 0.16 percent for non-export-base industries (Bartik 2017a, Table 5).

[3]This is the present value of the incentive package as a percent of the present value of value-added or wages as detailed in Bartik (2017a).

Within a state's export-base industries, incentives are not targeted by whether an industry is high-tech. As shown later, high-tech industries offer greater benefits to a local economy. Of 31 industries analyzed in Bartik (2017, Table 7), the most high-tech, as measured by high ratios of research and development spending to value-added, are chemicals manufacturing and computer manufacturing. Average incentives for chemicals and computers are 1.41 percent and 1.74 percent, ranking them twenty-fourth and tenth among the 31 industries.

In addition, incentives are not tightly tied to whether a state is short of jobs. The circles in Figure 1 each represent a state, with the circles proportional to population size. The horizontal axis shows the prime-age employment rate, while the vertical axis looks at a state's incentives as a ratio to value added of the firms receiving those incentives. States with a lower prime-age employment rate tend to have higher incentives, although as shown by the slope and the shaded area, the relationship is not strong or statistically significant. Large differences in incentives occur across nearby states for no clear economic reason. Compared to Illinois, Indiana's incentives are twice as high. As of 2015, Indiana's incentives averaged 2.68 percent of value-added for export-base industries; Illinois's were 1.35 percent, even though the two states have similar employment rates. Compared to North Carolina, South Carolina's incentives are twice as high; South Carolina had incentives of 2.39 percent, versus 0.93 percent in North Carolina, even though the two states have similar employment rates.

The ratio of state and local incentives to business value-added tripled from 1990 to 2015, as shown in Figure 2. Most of this increase occurred by 2001. (The jump from 2000 to 2001 was due to New York expanding its Empire Zone program statewide.) From 2001 to 2015, incentives were stable on average. Some high-incentive states, such as New York and Michigan, made cuts. Some low-incentive states, such as Wisconsin, made expansions.

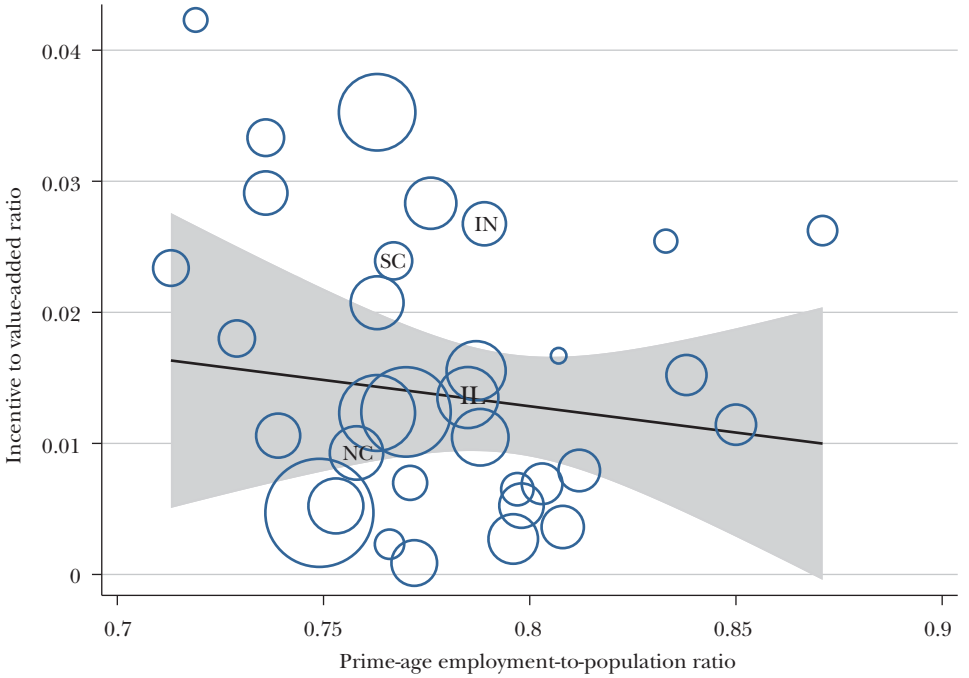
In sum, the current place-based jobs policies of state and local governments emphasize cash incentives to large corporations, without much targeting across industries or geographic areas. These patterns suggest some reforms that are likely to be desirable.

The Welfare Economics of Place-Based Jobs Policies: Some Preliminary Considerations

In the following four sections, I consider why policymakers might want to subsidize job creation in particular local labor markets, for either distributional or efficiency reasons. But first, I highlight special characteristics of local labor markets that are most relevant to this welfare economics analysis. These special characteristics include imperfect mobility and local job multipliers. I also highlight the issue of whether this policy intervention can be adequately addressed by state and local governments, or whether there is a need for federal involvement.

If mobility across local labor markets was perfect, the welfare economics analysis of place-based jobs policies would be simple, but uninteresting. Any benefits of local job creation would be fully capitalized into higher land values.

Figure 1
Incentive to Value-Added Ratio versus Employment Rate, Compared across States, as of 2015

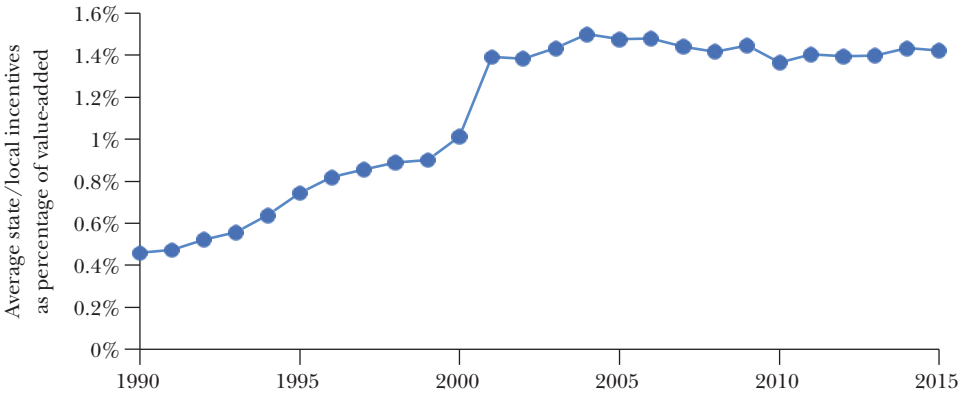


Note: Each circle represents one state. Vertical axis shows each state's 2015 ratio of the present value of incentives to the present value of value-added, for export-base industries as of 2015 from Bartik (2017b). Horizontal axis shows prime-age employment rate (ages 25–54) for each state as of 2015 and is from US Bureau of Labor Statistics (2015). Regression line and 95% confidence interval is shown based on weighted regression using 2015 state population (ages 25-54) as weights. Observations are on the 33 states with incentives data in Bartik (2017a). Circle size corresponds to state's 2015 population (ages 25-54). Four states referred to in text are initialed.

But people are hard to move. As Adam Smith (1776, Book I, Ch. 8) wrote, "a man is of sorts of luggage the most difficult to be transported." In cross-sectional data, about half of all Americans live within 30 miles of their birthplace (Zabek 2019). In panel data, about 55 percent of all Americans spend most of their career in their childhood metropolitan area (Bartik 2009). For college-educated workers, 40 percent spend their career in their childhood metro area. Percentages staying are not much lower in metro areas that are smaller or slower-growing.

These staying percentages reflect enhanced life satisfaction that people gain from the familiar people and places of their home area. As Roger Bolton (1992, p. 193) wrote, "The sense of place is an intangible, location-specific asset; it is capital . . . [People's] appreciation of [a sense of place] is evidenced by the one bit of evidence that ought to make economists notice: people are willing to pay for it." Estimates of the moving subsidy required for the median person to be indifferent to

Figure 2
Average State/Local Incentives as Percentage of Value-Added, 1990–2015



Source: From Bartik (2017a).

their current location and an otherwise similar location often exceed 100 percent of annual income (for a few recent examples, see A. Bartik 2018; Balgova 2018; Gregory 2017; T. Bartik 2019b includes further references). This required moving subsidy is 43 percent greater if a person's family is nearby (Kosar, Ransom, and van der Klauuw 2020).

But some people do move across local labor markets. Annual migration rates across metro areas exceed 2.5 percent, and five-year migration rates across metro areas or commuting zones exceed 10 percent (Molloy, Smith, and Wozniak 2011). For the thesis of this article, an important issue is the extent to which mobility leads the benefits of local job creation to be capitalized into higher land values rather than increasing local employment rates.

Local job creation policies should be cognizant of local job multipliers and how they vary across industries. If incentives are provided to non-export-base industries, the net effect on local job creation is negligible. Suppose an incentive encourages a local McDonald's to add jobs; this is likely to do little more than reduce sales and jobs at the local Burger King.

In export-base industries, local job multipliers are due to local supplier effects, worker demand effects, and agglomeration economies in high-tech industries. (High-tech agglomeration economies are discussed later.) Supplier-effect multipliers occur when a local expansion of some export-base firm leads to increased demand for local suppliers to that firm. Such supplier-effect multipliers will be higher for industries which use denser networks of local suppliers, like the auto industry. Worker-demand multipliers occur when workers in the expanding export-base firm, or its local suppliers, buy more locally produced or distributed retail products. Such worker-demand multipliers will be higher when worker wages are higher in the expanding export-base firm or its suppliers.

The local supplier effects and worker demand effects do not reflect any real externalities that would lead to market failures. These external effects on local suppliers and retailers are mediated through the market, via elastic responses to small changes in local prices facing suppliers and retailers. But if local job creation is sub-optimal for other reasons, to be discussed later, then the magnitude of local job multipliers determines how much net local job creation will occur if we induce job creation in an export-base firm. Higher local job multipliers increase the benefit-cost ratio for place-based jobs policies.

Average local job multipliers range from 1.3 to 1.7 (Bartik and Sotherland 2019). In other words, for every ten local jobs created in an export-base firm, another three to seven local jobs will be created. Multipliers will be higher for high-tech industries due to agglomeration economies (as will be discussed later).

In analyzing distributional or efficiency issues in local job creation, an important question is whether the appropriate policy response is best done at the state and local level or the federal level. Which level of government should be responsible is partially a technical economics issue: if state and local governments optimally correct for market failures of particular local labor markets, are there any remaining efficiency or equity issues at the national level that might rationalize federal intervention? Which level of government should "optimize" the pattern of local job creation is also a political issue, that is, at what level of government is wise action more likely? I will return to federal versus state/local responsibility throughout this article—in particular, in the penultimate section.

Distributional Rationale for Place-Based Jobs Policies

Before considering efficiency rationales for place-based jobs policies, I briefly consider the distributional rationale. Helping create jobs in distressed places could be argued to advance equity. Local labor markets with low employment to population ratios will tend to have lower incomes per capita. Place-based jobs policies, targeted at such distressed places, could be a mechanism for redistributing income to lower-income persons.

By itself, a concern for equity does not provide a strong case for place-based jobs policies. Place-based jobs policies would seem to be dominated by policies that target lower-income persons directly, such as "making the tax system more progressive or strengthening means-tested transfer programs" (Kline and Moretti 2014, p. 633). In the absence of market failures that allow local job creation to have benefits much greater than costs, place-based jobs policies are an unattractive way to help lower-income persons. To use Arthur Okun's (1975) "leaky bucket" metaphor, if we're trying to use a bucket to give water to the poor, the bucket of place-based jobs policy could be argued to be "leakier" than other buckets that might aid the poor.

One reason this bucket might be leaky: the benefits of place-based jobs policies may be diverted to upper-income groups. If there is extensive capitalization of local job creation into land values, then place-based jobs policies would have more of their benefits go to landowners, who tend to have higher incomes. If the costs of

subsidizing business to create jobs are large per job created, then place-based jobs policies would have more of their benefits going to owners of capital, who tend to have higher incomes. A later section will consider plausible magnitudes of capitalization and costs per job based on the empirical literature.

But as we will address, market failures can lead local job creation to have benefits greater than costs. Furthermore, these local benefits are distributed modestly progressively. Therefore, place-based jobs policies can potentially be a good way to deliver benefits to lower-income persons in distressed places. Maybe the "bucket" of place-based jobs policy somehow yields more water than you originally placed in the bucket. Equity concerns could rationalize pursuing place-based jobs policies more aggressively in distressed places and with federal support for such targeting.

Place-Based Jobs Policies and Market Failures, Part 1: Involuntary Unemployment

In the next three sections, I will address the role that place-based jobs policies can play in overcoming three different types of market failures that might cause local job growth to be inefficiently low: 1) involuntary unemployment; 2) agglomeration economies; and 3) problems in markets for various business inputs. The first two categories imply that local job growth may provide benefits external to the individual firm, and because firms do not internalize these external benefits, firms on their own may underprovide jobs. The last category analyzes how local job growth is impeded by inefficiencies in markets for business inputs. For more detailed discussion of place-related market failures, useful starting points are Bartik (1990), Kline and Moretti (2014), Glaeser and Gottlieb (2008), and Austin, Glaeser, and Summers (2018).

An understanding of these market failures gives clues to how to reform place-based jobs policies to increase benefits relative to costs. In particular, an important issue is whether external benefits of new jobs are asymmetric across places. After all, job gains in one place might easily be offset by job losses elsewhere. This is obviously true for a new plant which could have located elsewhere, but it is also true for local job growth due to start-ups or expansions. Greater national market share for firms in one place will mean losses of market share of firms elsewhere. Only if external benefits from jobs are asymmetric across places can redistributing jobs across places have net national gains. Identifying where jobs have more external benefits is useful advice for policymakers. Indeed, if places with greater external benefits don't do enough to attract jobs, perhaps the federal government should encourage further job redistribution.

Involuntary unemployment makes it likely that benefits from jobs are higher in distressed places. The social benefits from a higher employment rate include both the private benefits to individuals who otherwise would not be employed (that is, higher earnings minus the opportunity costs of their non-work time[4]), and external benefits like lower crime, benefits for family members, and local fiscal benefits.

[4]As discussed in Bartik (2012), one can visualize the social benefits of new jobs by thinking of how they create a vacancy chain terminated by increases in the local employment rate or population. Social

Even in nondistressed areas, the private benefits to the otherwise nonemployed are at least 40 percent of earnings, after subtracting opportunity costs of reduced non-work time (Mas and Pallais 2017; see also Haveman and Weimer 2015). In distressed places, both the benefits to those who otherwise would not be employed and the external social benefits are likely higher. In distressed areas, an increase in local jobs boosts total earnings for the area by more, by increasing local employment rates more. Simulations presented in Bartik (2015) suggest that in an initially low-unemployment area (4 percent unemployment rate), a 1 percent jobs boost will increase the local employment rate after 10 years by 0.20 percent. But in an initially high-unemployment area (7.1 percent unemployment), a similar-sized jobs boost will increase the employment rate after 10 years by 0.34 percent. Thus, the employment rate effect in a high-unemployment area is over two-thirds greater than in the low-unemployment area, and as a first-order approximation, local earnings per capita will also go up by over two-thirds more. Similarly, Austin, Glaeser, and Summers (2018, my calculations based on their Table 4, column 2) find that the impact of local job growth on prime-age male employment rates is 61 percent greater in areas with high nonemployment compared to areas with low nonemployment.

Why do local employment rates respond more to a job shock in distressed places? Consider what happens when a firm hires for new local jobs. The firm's hires come from three sources: 1) the local employed; 2) the local nonemployed; 3) in-migrants. Any hires from the local employed leads to local job vacancies, which will be filled in the same three ways. This job vacancy chain is only terminated when the original local jobs created lead to either additional jobs for the local nonemployed or for in-migrants. This makes sense because local employment is the product of the local employment rate and the local population, and thus a log percentage shock to local employment must equal the sum of the log percentage change in the local employment rate plus the log percentage change in the local population. All along this job vacancy chain, firms face choices about whether to hire the local nonemployed or in-migrants. Higher local nonemployment rates do mean that more local labor is readily available, as revealed by the empirical finding that employment rate effects of job shocks increase with greater local distress, which reflects changes in firms' decisions about who to hire.

Overall, job growth in more distressed areas will have greater private benefits for local workers, both because more local workers will obtain jobs in distressed areas and because the local workers who get jobs will value the jobs more relative to their opportunity costs. Furthermore, social benefits of job growth are also likely to be greater in distressed areas. Larger effects on local employment rates are likely to lead to greater crime reductions and greater improvements in outcomes for children. Fiscal benefits are also likely to be higher in distressed areas, both because of increased state and local tax revenue and because distressed areas are more likely to have excess capacity in infrastructure, which reduces needed spending for new infrastructure.

---

benefits can be measured by summing gains over the chain, or by the earnings gain minus the value of foregone non-work time for the otherwise nonemployed.

A common belief is that place-based jobs policy is a "zero-sum game," as any jobs gained in one place are argued to be lost to other places. What these greater benefits of job growth in distressed areas mean, however, is that reallocating jobs to distressed areas has net national benefits. Furthermore, this job reallocation might increase total national employment. This job reallocation moves labor demand to where effective labor supply responds more elastically without excess price pressures (Austin, Glaeser, and Summers 2018; Bartik 2001, Appendix 9).

On narrow grounds of efficiency, this market failure does not necessitate federal involvement. To maximize their own local benefits, state and local governments should provide the optimal subsidy for local job creation, equal to the marginal local benefits of job creation, which will increase with the local area's economic distress. Federal intervention could be rationalized on equity grounds to help distressed areas pay for the optimal job creation subsidy. I return to this topic in discussing policy reforms.

Place-Related Market Failures, Part II: Agglomeration Economies

External benefits of local job growth also occur because of agglomeration economies, which are productivity gains associated with city size or industry clustering (for an overview of this research, see the "Symposium on Productivity Advantages of Cities" in this issue). A larger scale of a city or industry cluster allows for better matching with more specialized suppliers and workers and for more knowledge in workers and firms to diffuse among local firms.

Because of agglomeration economies, job growth in one firm may have external benefits for other local firms. These external benefits are greater in certain industries, like high tech. However, research has not reached a consensus about how such external benefits are asymmetric across places. Will one more high-tech job produce greater agglomeration economies in Silicon Valley or in Detroit? We don't have a consensus answer.

Some research suggests that productivity may go up continuously with high-tech concentration so that high-tech job growth will have greater productivity spillovers in Silicon Valley than in lesser high-tech centers (Moretti 2019). Other research finds that local high-tech job multipliers are greater in areas whose high-tech concentration exceeds the national average but do not vary much among these above-average high-tech centers: for example, Silicon Valley compared to Minneapolis-St. Paul (Bartik and Sotherland 2019). These disparate results can be reconciled by allowing for congestion effects that high-tech clusters have on local costs due to higher housing prices and wages. The high-tech job multiplier in Bartik and Sotherland (2019) reflects both productivity spillovers and congestion effects.

This argument—that high-tech is similarly productive in many places—is an underlying assumption behind recent proposals for the federal government to encourage high-tech to diversify geographically. Jonathan Gruber and Simon Johnson (2019), in *Jump-Starting America*, propose that the federal government fund 20–30 high-tech centers. In a similar spirit, Atkinson, Muro, and Whiton (2019) offer

a more modest proposal for 8–10 high-tech centers. Gruber and Johnson (2019) argue that "a national strategy of developing new tech hubs [is] a cost-effective way to take advantage of local research spillovers and agglomeration." In their view, "conducting research in the existing small set of coastal locations is unambiguously a lot more expensive than doing the same in lower-cost locations elsewhere in the country." Geographic diversity of high-tech is desirable because "while there may or may not be some economic costs to redirecting public R&D towards new locations, there are unambiguous political gains." Their political argument will be considered further below.

Thus, the consensus is that there are efficiency gains from high-tech job growth in areas with above-average high-tech concentrations. But almost all high-tech jobs go to such areas anyway—mostly to a few high-tech clusters on the coasts (Johnson and Gruber 2019; Atkinson, Muro, and Whiton 2019). We lack consensus on which of the following two strategies would be better: diversifying high-tech away from coastal high-tech centers to other high-tech centers or concentrating high-tech more in coastal cities by reducing their costs through zoning and housing code reforms (Hsieh and Moretti 2019).

Despite this lack of consensus, agglomeration economies are relevant to place-based jobs policies because of how they affect multipliers. Local high-tech job multipliers probably average at least 1.9, compared to average job multipliers of 1.3 to 1.7 (Bartik and Sotherland 2019). In the many local labor markets with above-average high-tech clusters, local high-tech job multipliers may be as great as 2.5 to 3. Some studies estimate even larger high-tech multipliers (Moretti 2010). Place-based jobs policies can be more cost effective if they target high-tech firms with higher job multipliers.

One qualification to advocacy of high-tech targeting: it is unclear whether targeting high-tech has the same benefits for boosting local employment rates. Because many tech jobs are more skills-intensive, high-tech job growth by itself may not be well-matched to the job skills of the local nonemployed. However, many of the multiplier jobs will be less skills-intensive. The net effect on local employment rates of high-tech job growth, along with its multiplier jobs, needs more research.

Is federal intervention needed because of agglomeration effects? On efficiency grounds, no. Agglomeration gives state and local governments a reason for an optimal job creation subsidy varying by industry.

Federal intervention has been argued for on distributional grounds: it seems unfair to have most high-tech growth go to only a few coastal cities. Gruber and Johnson (2019) also make a political argument: more geographically diverse high-tech growth will increase political support for expanding federal R&D spending. The distributional case can be questioned: how will redistributing high-tech jobs from Silicon Valley to Minneapolis-St. Paul help low-income persons? The candidate high-tech centers in Gruber and Johnson (2019) or in Atkinson, Muro, and Whiton (2019) mostly seem relatively prosperous already. The political case also can be questioned. Why will rural Wisconsin residents, who already feel that the state capital of Madison gets unfair advantages (Cramer 2016), be more supportive of federal research and development if Madison is picked as a high-tech center?

Place-Relevant Market Failures, Part III: Public Services and Regulation Affecting Business Inputs

Place-based jobs policies can become more cost effective by being cognizant of market failures affecting business inputs. Business input costs may be affected by local public services and regulations. Here, I will focus on three local public policies affecting business inputs: customized business services, infrastructure, and land development. Suppose that per dollar spent on such programs, we can reduce business input costs by more than a dollar. In that case, such changes in public services and regulations can be more cost effective in promoting local job growth than tax incentives and other cash incentives. After all, providing cash to a business through tax incentives, at best, lowers business costs by one dollar per one dollar of government costs. These inefficiencies in business input provision do not necessitate federal intervention because state and local governments should be interested in creating jobs at lower costs.

Customized Business Services

Customized business services include manufacturing extension services for small manufacturers, customized job training provided by community colleges, small business development centers, and business incubators providing space, networking, and advice.

As one example, we already have a network of manufacturing extension services across the country funded by government and some private fees. These extension services provide advice to smaller manufacturers on adopting new technology, finding new markets, and other issues. Sometimes the advice, particularly if it is short-term, is directly provided for free by the extension office. Other times the advice is provided on a longer-term basis at a fee or by a network of high-quality advisors screened and recommended by the extension office. Some state extension services have links to local universities and community colleges and will provide referrals and some subsidies for manufacturers to receive consulting services from faculty.

Customized business services can be rationalized by several market failures. Many small firms lack adequate information, and these firms may also find it difficult to finance such services because of imperfect capital markets. The quality of information is hard to evaluate, and marginal costs of providing such information are low. Public agencies can provide higher-quality information or certify the quality of private or public consultants.

Customized job training includes both general and firm-specific training, which cannot be easily separated. Firms may underinvest in training if they fear worker exit. Such fears may be more acute for small firms. Smaller firms may also lack the scale to run training efficiently.

Quasi-experimental studies suggest productivity benefits for firms of at least five times the public costs for both manufacturing extension (Jarmin 1999) and customized job training (Holzer et al. 1993). The manufacturing extension evidence compares similar firms whose likelihood of receiving services varied with proximity

to the extension office. The customized training evidence compares similar manufacturers that applied at various times for a program that awarded grants on a first-come, first-serve basis. More recent survey evidence backs the effectiveness of both manufacturing extension (Robey et al. 2018) and customized job training (Hollenbeck 2008).

Public Infrastructure

In some cases, infrastructure can be a cost-effective way of promoting job growth in depressed regions. In one historical example, the Tennessee Valley Authority (TVA) emphasized infrastructure such as dams, highways, canals, and electrification. In research that compares the TVA region with similar regions that were unsuccessfully proposed for similar assistance, TVA is found to boost the region's manufacturing jobs by over 250,000 (Kline and Moretti 2013). Similarly, the Appalachian Regional Commission (ARC) focused two-thirds of its funding on highways. ARC highways had cumulative effects of increasing local jobs by 5.2 percent and increasing per capita incomes by 1.3 percent (Jaworski and Kitchens 2019). The present value of ARC highway investment is $85.1 billion, and the 1.3 percent boost in annual incomes is $13.4 billion.

Land Development

Land development in the United States is regulated by zoning, building codes, and tort litigation. Providing land that is appropriately zoned is widely viewed as having major effects on local job growth. Greater availability of industrially-zoned land affects development (Chapple 2014). Local economic developers believe making more land available is an effective development strategy. Promoting business development through "industrial parks" is a regular part of the local economic developer's tool kit.

Key Empirical Estimates Relevant to Place-Based Jobs Policies

Based on the above discussion, the strongest efficiency and equity rationale for place-based jobs policies is for creating jobs in distressed places. But is this rationale backed by estimates of benefits versus costs? Or are estimated benefits too small or too capitalized into land values? Or are estimated job-creation costs too high?

Before considering empirical estimates relevant to local job-creation's benefits and costs, I consider an oft-proposed alternative way of helping people in distressed areas: why not move people to jobs? Can a strategy to encourage out-migration from distressed areas work?

Moving People to Jobs Is an Ineffective Strategy

Encouraging out-migration from distressed places has two problems: 1) it's hard to get people to move; 2) out-migration does not help those left behind.

Even though annual and five-year migration rates among metro areas are large, getting additional people to move out of a distressed local labor market is costly.

Suppose we offer sizable subsidies for moving out of a distressed area, like a subsidy of $10,000. Subsidies in this range would increase out-migration rates by about 2 percentage points (Kennan and Walker 2011; A. Bartik 2018).

Even persistent local job loss has small out-migration effects. Autor, Dorn, and Hanson (2013, pp. 2141–42) "find no robust evidence that [Chinese trade–induced] shocks to local manufacturing lead to substantial changes in population." Going from the 25th to the 75th percentile of exposure to trade with China across commuting zones increases outmigration rates over a 10- to 14-year period by less than 1 percentage point, despite substantially lowering earnings (A. Bartik 2018).

Encouraging out-migration from distressed areas has another problem: it does little to re-equilibrate labor demand and supply in the distressed area. Migration shocks to local labor supply induce shocks to local labor demand of similar or larger size: for example, Beaudry, Green, and Sand (2018) and Howard (forthcoming) both find an elasticity of local jobs with respect to local population shocks of 1 or slightly higher; T. Bartik (2019b) provides a further review of the relevant literature. For every X percent of working-age people that leave Flint, Michigan, the jobs in Flint will go down by an amount similar to X percent, or perhaps more.

Why does migration affect local jobs? Migrants bring assets and transfer income, which affect local demand. Migrants demand housing, affecting construction. Migrants may be entrepreneurial, affecting business start-ups. Migration affects property values, which affect the local consumption of property owners (Howard forthcoming). Out-migration reverses all of these patterns.

A policy analysis of subsidizing out-migration from distressed communities should include all benefits and costs. Perhaps the out-migrant will get higher earnings. But if out-migrants choose not to move without the subsidy, their lost sense of place in their community may exceed the earnings boost they receive. If a person's unemployment imposes costs on their family, perhaps the family benefits of the move outweigh the individual's loss. But moving puts stress on children, which leads to behavioral problems, substance abuse, and poor mental health (Coley and Kull 2016; Oishi 2010). Subsidizing out-migration from distressed communities may often create as many problems as it solves.

In short, the problems caused by low employment rates in local labor markets cannot be substantially solved by encouraging people to leave for distant jobs. The main alternative is bringing jobs to people, or place-based jobs policies, which I turn to next. Do plausible empirical benefits of place-based jobs policies exceed plausible job-creation costs?

Benefits from Local Job-Creation: Is Everything Capitalized in the Long-Run?

Some economists have raised concerns that in the long run, benefits from job growth will be largely capitalized into land values (Glaeser and Gottlieb 2008; Marston 1985; Winnick 1966). In this view, higher employment rates and wages attract migrants, which will raise property values and push employment rates and real wages toward their prior equilibrium. Indeed, if capitalization were to dominate, local job growth would ultimately have regressive effects as property ownership is

concentrated in upper-income groups. But capitalization does not dominate; local job growth has sufficiently large benefits for local labor market outcomes, both in the short-run and more importantly in the long run.

An increase in local jobs has large short-run effects on employment rates: a short-run elasticity of about 0.6 (for reviews of 18 US studies, see T. Bartik 1993, 2015). Perhaps more surprising is that a one-time boost to the level of local jobs has large effects on local employment rates after 10 or more years. The consensus from the studies listed in Table 2 is a long-run elasticity of between 0.2 and 0.3.[5] These long-run employment-rate effects of a local job shock are qualitatively consistent with other research. More severe local recessions depress local employment rates for at least 25 years (Greenstone and Looney 2010) and real earnings per capita for at least 20 years (Stuart 2017). These long-run persistent effects of more severe local recessions hold in every recession from 1973 through the Great Recession (Hershbein and Stuart 2020).

Some historical examples reinforce these long-term effects. In areas where the government built a World War II manufacturing plant, manufacturing wages are higher even 50 years later (Garin 2018). Mississippi's "Balance Agriculture with Industry" program, a pioneering incentive program begun in 1936 (Freedman 2017), attracted northern manufacturing plants, mainly textile plants with a female workforce, by offering incentives of free land and buildings and property tax breaks. The BAWI program increased female labor force participation rates in the affected counties for at least 24 years until 1960. BAWI may also have increased male labor force participation. Effects persisted after most of the original plants had closed.

Human capital is a likely mechanism for these long-run effects. In the short run, local job growth increases residents' employment rates. This job experience improves their job skills, reduces crime and substance abuse, and increases self-confidence. Greater employment rates change social norms about work, and in some cases about women working (Freedman 2017). Even after migration has fully adjusted, residents' higher human capital allows higher employment rates.

Persistent local labor market effects are consistent with research on how labor demand shocks affect individuals. Worker displacement from jobs persistently lowers earnings (Davis and von Wachter 2011). Young workers entering the labor market during a recession suffer a long-run earnings penalty (Kahn 2010; Schwandt and von Wachter 2019). Locally severe recessions reduce residents' employment rates, even if the individual moves elsewhere (Yagan 2019).

These persistent employment rate effects of job growth imply large benefits per local job created. The long-run elasticity of 0.2 for the employment rate means that each job created raises earnings persistently by about 20 percent of average earnings per job. The present value of benefits per job will be many multiples of average earnings per job. Suppose all jobs paid $60,000 annually. Assume a fixed

[5] An exception is Blanchard and Katz (1992). However, their finding of zero long-run employment rate effects is not robust to alternative estimation approaches (T. Bartik 1993, 2015; Rowthorn and Glyn 2006). Amior and Manning (2018) argue that because data limitations forced Blanchard and Katz to include only two lags in all variables, long-run responses may be biased.

Table 2

Long-Run Elasticities of Local Employment to Population Ratio with Respect to Once and For All Local Employment Shock, US Studies

Study	*Nature of estimate*	*Long-run*	*Qualifications*	*Elasticity*
Bartik (2015)	Dynamic model, panel data on MSAs at annual frequency, 1979–2011	10-years	at 4.0% unemployment rate (UR)	0.20
		10-years	at 7.1% UR	0.34*
		10-years	at 10% UR	0.47*
Bartik (1991)	Reduced form regression of change in labor market outcomes on current and lagged annual job growth, annual panel data on MSAs, 1972–1986	8 years	OLS	0.23*
		8 years	2SLS using demand shock instruments	0.37*
Blanchard and Katz (1992)	Dynamic model with two lags, panel data on states with annual frequency, 1978–1990	8 years		0.07*
		17 years		0
Bartik (1993)	Same data as Blanchard and Katz but with lags in growth added	8 years		0.28*
		17 years		0.25*
Bound & Holzer (2000)	Decade change, 1980–1990, MSAs	10 years	High-school or less College or more	0.24* 0.12*
Partridge and Rickman (2006)	Dynamic model, annual panel data on states, 1970–1998	10 years	Preferred estimates	0.21*
			Alternative estimates	0.42*
Notowidigdo (forthcoming)	Decade changes, 1980–2000, panel data on MSAs	10 years	Mean effect	0.14*
Beaudry, Green, and Sand (2014)	"Decade" changes, 1970–2007, panel data on MSAs	10 years		0.24*
Amior and Manning (2018)	Decade changes, 1950–2010, panel data on commuting zones	10 years		0.30*

Notes: Elasticity is long-run elasticity of local employment to population ratio with respect to once-and-for-all shock to local job level. Asterisk indicates estimate is significantly different from zero at 5% significance level. Bartik (2015) figures extend that paper's results to slightly different low and middle unemployment rates, set at 10th and 90th percentile of local unemployment rates from 2016 ACS. Bartik (1991) OLS results from his Figures 4.2 and 4.3; 2SLS from his Table 4A4.2. Blanchard and Katz (1992) results are for a fixed shock to job level, which makes estimates comparable to other studies (T. Bartik, 1993, Table 2, Row 2). Bartik (1993) results are from his Table 2, Row 3, and add growth terms to Blanchard-Katz specification to optimize Akaike Information Criterion. Bound and Holzer (2000) results are IV results in their Table 3 for annual per capita hours worked. Partridge and Rickman (2006) comes from text discussion of their Figure 1, which implies statistical significance. Notowidigdo (forthcoming) results from his Table 2, column 3, using mean of predicted employment (obtained courtesy of the author), and then dividing resulting effect on the employment to population ratio in points by the mean ratio in his Table 1. Beaudry, Green, and Sand (2014) is from their Table 6, column 3. They use 2000 to 2007 as one of their "decade" changes. Amior and Manning (2018) results from their Table 2, column 4.

0.2 elasticity. Then each local job created increases employment rates sufficiently to increase earnings per capita by $12,000 (0.2 times $60,000). At a 3 percent discount rate, this earnings increase has a present value of $400,000. At a 5 percent discount rate, the present value is $240,000.

This calculation assumes that any opportunity costs of reduced non-work time, due to higher employment rates, are at least offset by social benefits like lower crime and substance abuse and higher fiscal benefits. This rough calculation also ignores

the higher short-run elasticity, as well as the possibility that this long-run elasticity might eventually depreciate. But more refined calculations are also likely to yield benefits per local job created in the hundreds of thousands of dollars.

Do Employment Rate Effects Outweigh Capitalization's Regressive Effects?

Using plausible empirical magnitudes, a demand shock to the level of local jobs affects the local income distribution progressively. The elasticity of housing prices with respect to a local job shock is 0.4–0.5 (T. Bartik 1991; see also Saiz 2010). But a shock to the local job level results in only a one-time gain for property owners. In contrast, a local job shock increases local employment rates for many years—an elasticity of 0.6 in the short run and 0.2 to 0.3 in the long run. Real wage rates also rise (T. Bartik 2015). Under plausible discount rates, the present value of these continuing increases in local per capita earnings will exceed the one-time property-value gain by a ratio of over 3 to 1 (T. Bartik 1994, 2005, 2018a).[6] Real earnings effects of local job growth are modestly progressive (Bartik 1994) in part because higher employment rates help those otherwise not employed who tend to have lower incomes.

Because earnings effects dominate property-value effects, and earnings effects are progressive, the overall effects of local job growth are progressive. Table 3 summarizes some illustrative estimates. Progressive impacts are modest. Percentage effects on the lowest-income quintile are 2.4 times percentage effects on the highest-income quintile. Because the highest-income quintile has 10 times the average income of the lowest-income quintile, however, the dollar effects of local job growth on incomes are higher for higher-income groups.

As a result, how we finance local job growth makes a big difference to whether such policies have a progressive effect. If the financing is by cutting highly progressive public spending, such as welfare (Bartik 1994) or public schools (T. Bartik 2018a), the net impact is likely to be regressive. Also, the progressivity of local job growth may vary with policy. If a firm receiving incentives hires more of the local non-employed, the employment rate effect of the new jobs will be greater. A better local workforce system might encourage hiring of the local non-employed. Capitalization is greater in local areas with more housing supply restrictions (Saiz 2010). However, even in areas with the most-restricted housing supply, the long-run earnings effects of local job growth have a greater present value than the increased housing values (T. Bartik 2018a). The match of the jobs created with local skills may also matter, but there is little empirical evidence on this point: for some simulations, see Persky, Felsenstein, and Carlson (2004).

Costs of Creating Local Jobs: High for Cash Incentives, Lower for Policies Affecting Business Inputs

[6]Why is the ratio of earnings effects to property value effects so high? In T. Bartik (2018a), locally-owned property is 3.5 times earnings, and the property value elasticity is 0.45, so the property value effect is equivalent to a one-year-only earnings elasticity of 1.6. If the earnings elasticity starts at 0.6 and declines to 0.3, and real wages also increase, it is unsurprising that labor market benefits exceed property value effects.

Even if local job creation can have high benefits per job created, do these benefits outweigh the costs of job creation? The answer is: maybe for cash incentives but more clearly for policies that enhance business inputs.

Cash incentives to encourage local job creation have high costs per job created because it takes a lot of cash to tip a business location or expansion decision. The empirical literature suggests that a cash incentive with a present value of 1 percent of value-added—slightly below the average for state and local incentives—will tip, at most, 10 percent of the location or expansion decisions of businesses awarded such incentives (for a review of the relevant literature, see Bartik 2020). At least 90 percent of the time, the business receiving the incentive would have made the same location or expansion decision, even without the incentive.

Why such modest effects? A cash incentive with a present value of 1 percent of value-added is equivalent to a perpetual 2.2 percent wage subsidy. Differences across locations in other costs—wages, labor productivity, real estate, proximity to markets or suppliers—will often outweigh a 2 percent wage subsidy. An anecdote: One CEO told me that his decision about where to locate a particular new facility had been determined by the availability in that city of an empty factory. The empty factory allowed the new facility to get into production quickly. The incentive his business received had no effect on the location decision.

Such modest effects of incentives imply that the cost per job created will be high: at least $100,000 and probably more. In 2019 dollars, value-added per full-time-equivalent worker in export-base industries is $177,000 (Bartik 2017, Table 3). According to Poterba and Summers (1995), firms use a 12 percent real discount rate in making investment decisions. At a 12 percent real discount rate, the present value of value-added per full-time-equivalent in export-base industries is $1.65 million ($177,000/0.12). Suppose an upfront incentive is provided. Based on the research, an incentive of 1 percent of the present value of value-added will tip 10 percent of such location decisions. This implies a cost per job created in incented firms of $165,000 ($1.65 million times 0.01/0.10). If the local job multiplier is 1.5, then the cost per job created, including multiplier jobs, will be around $110,000 ($165,000/1.5).

But a totally upfront incentive is not the typical incentive package. The typical incentive package continues at a high level for at least 10 years (Bartik 2017, Table 22). States prefer drawn-out incentives to up-front incentives for two reasons. First, up-front incentives pose a greater problem of clawing back the incentive if the jobs go away. Second, governors prefer to defer incentive costs to their successors. More drawn-out incentive payments will have decreased cost-effectiveness relative to their social costs as long as the social discount rate is lower than the discount rate used by firms in making location decisions. Using the typical drawn-out incentive structure in the U.S., a social discount rate of 3 percent, and a multiplier of 1.5, the model described in T. Bartik (2018a) finds a cost per job created of $196,000.

As mentioned, the benefits per job created may also be in the hundreds of thousands of dollars. Therefore, even costly incentives may have benefits greater than costs. But often the benefit-cost ratio for incentives will not be much greater than one. For example, one study finds that when incentives are targeted at

Table 3

Percentage Effects of a 1% Local Job Shock on Income Due to Earnings Effects and Property Value Effects

Type of effect	*Average percent effects on income for:*		
	All households	*Lowest income quintile*	*Highest-income quintile*
Earnings effects	0.18%	0.41%	0.11%
Property value effects	0.05%	0.03%	0.07%
Sum of earnings effects and property value effects	0.23%	0.44%	0.18%

Note: Based on simulation model of T. Bartik (2018a). This model assumes short-run and long-run employment rate elasticities of 0.6 and 0.25, and housing price elasticity of 0.45. The distribution of earnings effects by quintile is from Bartik (1994). This version of the T. Bartik (2018a) model assumes fixed 1% once and for all shock to local job level that occurs at no incentive cost. Effects calculated over 20 years and use a 3% social discount rate. Present value of earnings effects and property value effects, after state/local taxes, are calculated as percentage of present value of average market plus transfer income of each group. Based on Congressional Budget Office (2016), the baseline income shares of the lowest- and highest-income quintiles are 5.1% and 52.0%.

non-distressed areas and have an average multiplier, the estimated benefit-cost ratio is 1.5 (Bartik 2019a). This leaves little room for plausible alternative assumptions.

Incentives are more likely to pay off if multipliers are higher or if incentives are targeted at distressed areas. Higher multipliers reduce costs per total local job created. Targeting at distressed areas increases benefits per job created.

Costs per job created can be much lower for policies affecting business inputs. As discussed, some studies suggest that customized public services to business can lower business costs by about five times their costs to the government. If so, such services can have a cost per job created that is much lower than cash incentives. Using the model in T. Bartik (2018a), these customized services have a cost per job created of around $34,000. Based on surveys of firms, manufacturing extension and customized job training both have a cost per job created of under $15,000 (Ehlen 2001; Hollenbeck 2008).

Costs per job created for infrastructure programs are also lower than for cash incentives. Consider the Tennessee Valley Authority, which improved local electrification as well as providing other services. TVA cost a little around $30 billion (in 2019 dollars). As mentioned, TVA is estimated to increase manufacturing jobs by a little over 250,000 (Kline and Moretti 2013). In this large multi-state area, the multiplier effect of this job creation would be higher, plausibly at least 2 (Bartik and Sotherland 2019), so the total job creation would be 500,000. The cost per job would then be around $60,000. A more refined calculation, which corrects for the timing of TVA spending versus job creation, finds that TVA's present value cost per job created is around $77,000 (for more details, see the Data Repository for this article).

A lower cost per job created also can be achieved by making more land available for business development. For example, case studies suggest that cleaning up "brownfields"—older industrial sites with environmental contamination—may have a cost per job created of $13,000 (Paull 2008).

All these estimates have the inherent uncertainty of any research finding. But another source of uncertainty is that effects of public services to business, infrastructure, and land development practices will probably vary with both program quality and local context. The cost-effectiveness of manufacturing extension and customized job training depend on the quality of the local programs as well as on the characteristics of local businesses. The success of infrastructure investments in rural electrification by the Tennessee Valley Authority does not mean that the proverbial "bridge to nowhere" is a good idea. Whether a brownfield project or industrial park succeeds in attracting industry depends very much on "location, location, location"—as well as on the local economy.

Improving Place-Based Jobs Policies

Based on this analysis and the empirical literature, how might place-based jobs policy be improved? Here I list six possible reforms. At the outset, I emphasize that these reforms include evaluation because we need to know whether a particular program is working in a particular place.

First, place-based jobs policies should be more geographically targeted to distressed places. The benefits of more jobs are at least 60 percent greater in distressed places than in booming places. But our current incentive system does not significantly favor distressed places.

Second, place-based jobs policies should be more targeted at high-multiplier industries, such as high-tech industries. Governors may claim they want to build the future economy, but state and local governments in practice do not target high-tech. One caveat: high-tech targeting should consider how to increase the access of current residents to these jobs. One model is Virginia's recent offer for Amazon's "Headquarters II." Virginia's offer included a new Virginia Tech campus in northern Virginia and increased funding at state colleges for tech-related programs. These education programs increased the odds that Virginia residents would fill Amazon's jobs.

Third, incentives should not disproportionately favor large firms, especially given the renewed concern in economics over excess market power in product markets and labor markets (Azar, Marinescu, and Steinbuam 2017; Gutiérrez and Phillippon 2017).

Fourth, place-based jobs policies should put more emphasis on enhancing business inputs. Customized business services, infrastructure, and land development services have the potential to be more cost effective than incentives as ways to increase local jobs and earnings.

Fifth, place-based policies should be a coordinated package of policies attuned to local conditions. One area may need more infrastructure; another, training; and still another, better land development processes. Place-based policies are complementary. If the local nonemployed are more skilled, job growth increases employment rates more. If more jobs are available, it is easier to design effective training programs. Business inputs are complementary—boosting infrastructure helps growth more if the local economy also has customized business services.

Sixth, place-based jobs policies should be evaluated better. Random assignment of firms or areas receiving assistance is hard to implement. However, better evaluation is attainable by awarding assistance using quantifiable selection criteria because this method allows for regression discontinuity techniques comparing firms or areas that just made or missed the cutoff for receiving services. For example, support for small business services can be evaluated better if some quantitative scoring system is used to decide which firms receive services. Area strategies can be evaluated better if higher units of government select distressed areas using a quantitative cutoff. Some improvements have been made in transparency and evaluation for incentives (Pew Charitable Trusts 2017).

Taking these themes together, it seems plausible that an appropriate scale of place-based jobs policies could have lower costs than current policies. For a back-of-the-envelope estimate of the needed number of jobs, if we take all commuting zones in the bottom quintile of employment rates and ask how many jobs would be needed to bring their employment rates to the median, assuming an elasticity of 0.25 for the employment rate, the needed job creation is 6 million jobs. Based on the previous discussion, a job-creation strategy that focused on enhancing business inputs might create jobs at $50,000 per job. The total cost to generate or reallocate 6 million jobs to distressed areas would then be $300 billion. If pursued over a 10-year period, the cost would be $30 billion annually. This is half of current annual costs for place-based jobs policies of $60 billion. This less costly strategy would do more to enhance both efficiency and equity by bringing up the bottom quintile of commuting zones to the national median employment rate.

Is Federal Intervention Needed?

An efficient place-based jobs program is feasible without federal intervention. State and local governments should find it in their residents' interests to pursue job creation more aggressively in distressed places. All places can improve cost-effectiveness by better targeting of incentives and by diverting resources from incentives to enhancing business inputs. But the ideal isn't happening. The political temptation is strong both in distressed and nondistressed places for state and local governments to devote lots of money to incentives to large firms, without much industry targeting. Voters are more likely to support incumbent governors who offer high-profile incentives (Jensen and Malesky 2018). Should the federal government intervene?

Any intervention should be realistic about federal capacities and local needs. First, the federal government already has a lot on its plate. Federal knowledge is limited about what works best to advance economic development in particular places.

Second, any meaningful system of federalism must allow state and local governments considerable freedom to pursue their own goals for economic development in their own ways. Given the diversity of local needs and our still-evolving

knowledge about what works in economic development, there are advantages to local experimentation.

Therefore, federal intervention in place-based jobs policies should be limited to promoting clear national interests. One national interest is limiting unfair advantages for larger firms. Large incentives for the largest companies go against the national interest in limiting large firms' market power. Another national interest is helping people in distressed places. As mentioned, distressed places have good reason, on their own, to aggressively promote job creation using their own resources. But distressed places' resources are more limited. The national interest in equity would be advanced by helping pay for job-creation programs in distressed places.

How could the federal government promote these national interests while respecting local needs? I offer below some suggestions, for both limiting incentives and helping distressed places.

Capping Incentives for Large Firms

Federal limits on incentives have been previously proposed. Incentives could be subjected to extra federal taxes (Burstein and Rolnick 1995) or penalized by reducing federal grants (Chatterji 2018; LeRoy 2012).

But such regulation of incentives raises two problems. First, there is the problem of how to define an "incentive" versus normal tax policy and how to administer any federal restrictions, given the thousands of incentives handed out each year. Second, such incentive regulation might imply federal micro-management of the state and local business tax system and state and local economic development policy.

To overcome these problems, the federal government could pursue a more modest goal: capping incentives to the largest firms. The federal government could enforce such limits by an outright ban, an extra tax, or withholding federal grants. What would be limited is discretionary incentives to firms with more than, say, 10,000 employees that exceeded some percentage of the investment or annual payroll in the new jobs created. States would remain free to provide incentives to all firms, all export-base firms, or all manufacturing firms. States could decide how much they want job growth, and in large measure how best to pursue it. What would be restricted is providing extra incentives to large firms simply because they are already large.

Targeting incentive restrictions on larger firms reduces the federal government's administrative challenges in managing incentive restrictions. As of 2016, only 1,491 US firms had more than 10,000 employees (Longitudinal Business Database, US Census Bureau 2018). Such firms are 28 percent of business employment but receive over half of state and local incentives. Restricting incentives to such large firms might reduce incentives by one-third.

This proposed incentive restriction has a precedent in European Union rules on regional state aid (LeRoy and Thomas 2019). The European Union limits the magnitude of such state aid, with stricter limits on larger firms. The European Union also makes the limits stricter in less-distressed areas.

Helping Distressed Areas with Flexible Block Grants

The federal government could help distressed areas by providing economic development grants. Many expanded spending programs for distressed places have been recently proposed: infrastructure spending (Center for American Progress 2018; Smith 2018); public jobs programs or nonprofit jobs programs (Center for American Progress 2018; Neumark 2018); employer subsidies for job creation or hiring, with some targeting on disadvantaged workers (Glaeser, Summers, and Austin 2018; Neumark 2018); manufacturing extension (Baron, Kantor, and Whalley 2018; Bartik 2010); customized job training (Austin, Glaeser, and Summers 2018; Bartik 2010); and help for small business or entrepreneurs (Chatterji 2018)

However, helping distressed places with only one specific program has problems. A single-program approach assumes that researchers know the "one best program." We don't. A single-program approach assumes that all distressed places need the same program. They don't. A single program does not allow for the synergy from simultaneously pursuing multiple programs. For example, a wage subsidy to encourage the hiring of disadvantaged workers will create more jobs and have smaller displacement effects if it is combined with customized business services that promote local job creation in businesses with high multipliers. A better federal approach to helping distressed areas is to provide a flexible block grant with many allowable uses, which can then be attuned to local needs.

This federal block grant program can also promote better evaluation. As noted earlier, the federal government could design its selection of distressed areas so that distressed and nondistressed areas can be compared for evaluation purposes.

Conclusion

Places matter for policy because places matter to people. Our knowledge about local labor markets should inform how we help distressed places. Our most important knowledge can be summarized in two numbers discussed earlier in this essay: 1.0 and 0.2. A shock to local population has an elasticity of about 1.0 in its effects on local jobs. Therefore, moving people from distressed to nondistressed places does not help restore local labor market equilibrium. A shock to local jobs of 1 percent increases the local employment rate in the long run by at least 0.2 percent. This allows place-based jobs policy to have large long-run benefits that are distributed progressively.

Place-based jobs policy needs additional research. As this essay has argued, reformed place-based jobs policies can have higher benefit-cost ratios than current policies. The targeting of distressed places could be improved with a more extensive research basis for defining distressed places and identifying which programs are most cost effective in different places. But the existing evidence clearly shows that adding jobs in distressed places offers both private and social benefits.

■ *I appreciate helpful suggestions from Enrico Moretti, Heidi Williams, Gordon Hanson, Timothy Taylor, Ben Jones, Jeff Chapman, Mark Robyn, George Erickcek, Adam Ozimek, Brian Asquith, Richard Florida, Sue Helper, Mark Muro, Brad Hershbein, Michelle Miller-Adams, Sue Houseman, and Alex Bartik. I also appreciate the assistance of Nathan Sotherland, Shane Reed, Ken Kline, Alexandra Szczupak and Claire Black.*

References

Amior, Michael, and Alan Manning. 2018. "The Persistence of Local Joblessness." *American Economic Review* 108 (7): 1942–70.

Atkinson, Robert D., Mark Muro, and Jacob Whiton. 2019. *The Case for Growth Centers: How to Spread Tech Innovation across America.* Washington, DC: Brookings Institution.

Austin, Benjamin, Edward Glaeser, and Lawrence H. Summers. 2018. "Saving the Heartland: Place-based Policies in 21st Century America." *Brookings Papers on Economic Activity* Spring (2018).

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review* 103 (6): 2121–68.

Autor, David H., David Dorn, and Gordon H. Hanson. 2018. "When Work Disappears: Manufacturing Decline and the Falling Marriage-Market Value of Young Men." Casio Working Paper Series 7010, CESifo Group Munich.

Azar, José, Ioana Marinescu, and Marshall I. Steinbaum. 2017. "Labor Market Concentration." NBER Working Paper 24147.

Balgova, Maria. 2018. "Why Don't Less Educated Workers Move? The Role of Job Search in Migration Decisions". Job market paper, Department of Economics, University of Oxford.

Baron, E. Jason, Shawn Kantor, and Alexander Whalley. 2018. "Extending the Reach of Research Universities: A Proposal for Productivity Growth in Lagging Communities." In *Place-Based Policies for Shared Economic Growth*, edited by Jay Shambaugh and Ryan Nunn, 157–84. Washington, DC: Brookings.

Bartik, Alexander W. 2018. "Moving Costs and Worker Adjustment to Changes in Labor Demand: Evidence from Longitudinal Census Data." Working paper, Department of Economics, University of Illinois at Urbana-Champaign.

Bartik, Timothy J. 1990. "The Market Failure Approach to Regional Economic Development Policy." *Economic Development Quarterly* 4 (4): 361–70.

Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Bartik, Timothy J. 1993. "Who Benefits from Local Job Growth: Migrants or the Original Residents?" *Regional Studies* 27 (4): 297–311.

Bartik, Timothy J. 1994. "The Effects of Metropolitan Job Growth on the Size Distribution of Family Income." *Journal of Regional Science* 34 (4): 483–501.

Bartik, Timothy J. 2001. *Jobs for the Poor: Can Labor Demand Policies Help?* New York: Russell Sage Foundation.

Bartik, Timothy J. 2005. "Solving the Problems of Economic Development Incentives." *Growth and Change* 36 (2): 139–66.

Bartik, Timothy J. 2009. "What Proportion of Children Stay in the Same Location as Adults, and How Does This Vary across Location and Groups?" Upjohn Institute Working Paper 09-145. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. https://doi.org/10.17848/wp09-145.

Bartik, Timothy J. 2010. "Bringing Jobs to People: How Federal Policy Can Target Job Creation for Economically Distressed Areas." The Hamilton Project Discussion Paper. Washington, DC: Brookings Institution.

Bartik, Timothy J. 2012. "Including Jobs in Benefit-Cost Analysis." *Annual Review of Resource Economics* 4 (1): 55–73.

Bartik, Timothy J. 2015. "How Effects of Local Labor Demand Shocks Vary with the Initial Local Unemployment Rate." *Growth and Change* 46 (4): 529–57.

Bartik, Timothy J. 2017a. "A New Panel Database on Business Incentives for Economic Development Offered by State and Local Governments in the United States." Report prepared for the Pew Charitable Trusts.

Bartik, Timothy J. 2017b. "The PDIT Database." Kalamazoo, MI: The W.E. Upjohn Institute for Employment Research. https://www.upjohn.org/bied/database.php (accessed May 27, 2020).

Bartik, Timothy J. 2018a. "Who Benefits from Economic Development Incentives? How Incentive Effects on Local Incomes and the Income Distribution Vary with Different Assumptions about Incentive Policy and the Local Economy." Upjohn Institute Technical Report 18-034. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. http://doi.org/10.17848/tr18-034.

Bartik, Timothy J. 2018b. "What Works to Help Manufacturing-Intensive Local Economies?" Upjohn Institute Technical Report 18-035. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. https://doi.org/10.17848/tr18-035.

Bartik, Timothy J. 2019a. *Making Sense of Incentives: Taming Business Incentives to Promote Prosperity*. Kalamazoo, MI: Upjohn Institute for Employment Research.

Bartik, Timothy J. 2019b. "Should Place-Based Jobs Policies Be Used to Help Distressed Communities?" Upjohn Institute Working Paper 19-308. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. https://doi.org/10.17848/wp19-308.

Bartik, Timothy J. 2020. "Introduction to Special Issue: Learning More About Incentives." *Economic Development Quarterly*. https://doi.org/10.1177/0891242420916033.

Bartik, Timothy J., and Nathan Sotherland. 2019. "Local Job Multipliers in the U.S.: Variation with Local Characteristics and with High-Tech Shocks". Upjohn Institute Working Paper 19–301. http://dx.doi.org/10.2139/ssrn.3379722.

Bastian, Jacob, and Katherine Michelmore. 2018. "The Long-Term Impact of the Earned Income Tax Credit on Children's Education and Employment Outcomes." *Journal of Labor Economics* 36 (4): 1127–63.

Beaudry, Paul, David A. Green, and Benjamin M. Sand. 2014. "Spatial Equilibrium with Unemployment and Wage Bargaining: Theory and Estimation." *Journal of Urban Economics* 79 (1): 2–19.

Beaudry, Paul, David A. Green, and Benjamin M. Sand. 2018. "In Search of Labor Demand." *American Economic Review* 108 (9): 2714–57.

Blanchard, Olivier Jean, and Lawrence F. Katz. 1992. "Regional Evolutions." *Brookings Papers on Economic Activity* 1: 1–75.

Bolton, Roger. 1992. "'Place Prosperity vs. People Prosperity' Revisited: An Old Issue with a New Angle." *Urban Studies* 29 (2): 185–203.

Bound, John, and Harry J. Holzer. 2000. "Demand Shifts, Population Adjustments, and Labor Market Outcomes during the 1980s." *Journal of Labor Economics* 18 (1): 20–54.

Burstein, Melvin L., and Arthur J. Rolnick. "Congress Should End the Economic War among the States**."** *Federal Research Bank of Minneapolis,* January 1995, https://www.minneapolisfed.org/article/1995/congress-should-end-the-economic-war-among-the-states.

Center for American Progress. "Blueprint for the 21st Century: A Plan for Better Jobs and Stronger Communities." *The Center for American Progress,* May 14, 2018, https://www.americanprogress.org/issues/economy/reports/2018/05/14/450856/blueprint-21st-century/.

Council for Community and Economic Research. 2017. *State Economic Development Program Expenditure Trends, FY2014-FY2016*. Report, C2ER.

Council for Community and Economic Research. 2018. *State Economic Development Program Expenditures Database*. Online database.

Chapple, Karen. 2014. "The Highest and Best Use? Urban Industrial Land and Job Creation." *Economic Development Quarterly* 28 (4): 300–313.

Charles, Kerwin Kofi, Erik Hurst, and Mariel Schwartz. 2018. "The Transformation of Manufacturing and the Decline in U.S. Employment." *NBER Macroeconomics Annual* 33 (2018): 307-72

Chatterji, Aaron K. 2018. "The Main Street Fund: Investing in an Entrepreneurial Economy." The Hamilton Project Policy Proposal 2018-09. Washington, DC: Brookings Institution.

Coley, Rebekah Levine, and Melissa Kull. 2016. "Cumulative, Timing-Specific, and Interactive Models of Residential Mobility and Children's Cognitive and Psychosocial Skills." *Child Development* 87 (4): 1204–20.

Congressional Budget Office (CBO). 2016. "The Distribution of Household Income and Federal Taxes, 2013." Washington, DC: Congressional Budget Office.

Cramer, Katherine. 2016. *The Politics of Resentment: Rural Consciousness in Wisconsin and the Rise of Scott*

Walker. Chicago: University of Chicago Press.

Davis, Steven J., and Till von Wachter. 2011. "Recessions and the Costs of Job Loss." *Brookings Papers on Economic Activity* Fall(2011): 1–72.

Diette, Timothy M., Arthur H. Goldsmith, Darrick Hamilton, and William Darity. 2018. "Race, Unemployment, and Mental Health in the USA: What Can We Infer about the Psychological Cost of the Great Recession across Racial Groups?" *Journal of Economics, Race, and Policy* 1: 75–91.

Ehlen, Mark A. 2001. "The Economic Impact of Manufacturing Extension Centers." *Economic Development Quarterly* 15 (1): 36–44.

Freedman, Matthew. 2017. "Persistence in Industrial Policy Impacts: Evidence from Depression-Era Mississippi." *Journal of Urban Economics* 102: 34–51.

GAO. 2012a. *Limited Information on the Use and Effectiveness of Tax Expenditures Could Be Mitigated through Congressional Attention*. GAO-12-262. Washington, DC: GAO.

GAO. 2012b. *Economic Development: Efficiency and Effectiveness of Fragmented Programs Are Unclear*. GAO-12-553T. Washington, DC: GAO.

Garin, Andrew. "Essays on the Economics of Labor Demand and Policy Incidence." Doctoral diss., Harvard University, 2018 . http://nrs.harvard.edu/urn-3:HUL.InstRepos:41127168.

Glaeser, Edward L., and Joshua D. Gottlieb. 2008. "The Economics of Place-Making Policy." NBER Working Paper 14373.

Glaeser, Edward L., Lawrence H. Summers, and Ben Austin. "A Rescue Plan for a Jobs Crisis in the Heartland." *The New York Times*, May 24, 2018.

Greenstone, Michael, and Adam Looney. 2010. "An Economic Strategy to Renew American Communities." The Hamilton Project Strategy Paper. Washington, DC: Brookings Institution.

Gregory, Jesse. 2017. "The Impact of Post-Katrina Rebuilding Grants on the Resettlement Choices of New Orleans Homeowners." Working paper, University of Wisconsin.

Gruber, Jonathan, and Simon Johnson. 2019. *Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream*. New York: PublicAffairs.

Gutiérrez, Germán, and Thomas Philippon. 2017. "Investment-less Growth: An Empirical Investigation." NBER Working Paper 22897.

Haveman, Robert H., and David L. Weimer. 2015. "Public Policy Induced Changes in Employment: Valuation Issues for Benefit-Cost Analysis." *Journal of Benefit Cost Analysis* 6 (1): 112–53.

Helliwell, John F., and Haifang Huang. 2014. "New Measures of the Costs of Unemployment: Evidence from the Subjective Well-Being of 3.3 Million Americans." *Economic Inquiry* 52 (4): 1485–1502.

Hershbein, Brad, and Bryan A. Stuart. 2020. "Recessions and Local Labor Market Hysteresis". Upjohn Institute Working Paper 20-325. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Hollenbeck, Kevin. 2008. "Is There a Role for Public Support of Incumbent Worker On-the-Job Training?" Upjohn Institute Working Paper 08-138. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Hollenbeck, Kevin. 2013. "Public Financing of Workforce Services for Incumbent Workers?" Working Paper, School of Public Policy, University of Maryland.

Holzer, Harry J., Richard N. Block, Marcus Cheatham, and Jack H. Knott. 1993. "Are Training Subsidies for Firms Effective? The Michigan Experience." *Industrial and Labor Relations Review* 46 (4): 625–36.

Howard, Greg. Forthcoming. "The Migration Accelerator: Labor Mobility, Housing, and Demand." *American Economic Journal: Macroeconomics*.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics* 11 (2): 1–39.

Jarmin, Ronald S. 1999. "Evaluating the Impact of Manufacturing Extension on Productivity Growth." *Journal of Policy Analysis and Management* 18 (1): 99–119.

Jaworski, Taylor, and Carl T. Kitchens. 2019. "National Policy for Regional Development: Evidence from Appalachian Highways." *Review of Economics and Statistics* 101 (5): 777–90.

Jensen, Nathan M., and Edmund J. Malesky. 2018. *Incentives to Pander: How Politicians Use Corporate Welfare for Political Gain*. Cambridge: Cambridge University Press.

Kahn, Lisa B. 2010. "The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy." *Labour Economics* 17 (2): 303–16.

Kennan, John, and James R. Walker. 2011. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79 (1): 211–51.

Kimbo, Tatiana, and Richard Phillips. "How Opportunity Zones Benefit Investors and Promote Displacement." *Institute on Taxation and Economic Policy*, August 10, 2018. https://itep.org/how-opportunity-zones-benefit-investors-and-promote-displacement/.

Kline, Patrick, and Enrico Moretti. 2013. "Local Economic Development, Agglomeration Economies,

and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *The Quarterly Journal of Economics* 129 (1): 275–331.

Kline, Patrick, and Enrico Moretti. 2014. "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." *Annual Review of Economics* 6: 629–62.

Kosar, Gizem, Tyler Ransom, and Wilbert van der Klaauw. 2020. "Understanding Migration Aversion Using Elicited Counterfactual Choice Probabilities." IZA Discussion Papers 12271. Bonn, Germany: IZA Institute of Labor Economics.

LeRoy, Greg. "What the Country Needs Now, Mr. President." *Planning,* November, 2012. https://www.goodjobsfirst.org/sites/default/files/docs/pdf/planning_op-ed_nov2012.pdf.

LeRoy, Greg, and Kenneth Thomas "Lessons for the U.S.: How the EU Controls Bidding Wars for Jobs and Investment." *Shelterforce,* June 17, 2019. https://shelterforce.org/2019/06/17/lessons-for-the-u-s-how-the-eu-controls-bidding-wars-for-jobs-and-investment/.

Marston, Stephen T. 1985. "Two Views of the Geographic Distribution of Unemployment." *The Quarterly Journal of Economics* 100 (1): 57–79.

Mas, Alexandre, and Amanda Pallais. 2017. "Labor Supply and the Value of Non-Work Time: Experimental Estimates from the Field." NBER Working Paper 23906.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic Perspectives* 25 (3): 173–96.

Moretti, Enrico. 2010. "Local Multipliers." *American Economic Review: Papers & Proceedings* 100(May): 373–77.

Moretti, Enrico. 2019. "The Effect of High-Tech Clusters on the Productivity of Top Inventors". NBER Working Paper 26270.

Neumark, David. 2018. "Rebuilding Communities Job Subsidies." In *Place-Based Policies for Shared Economic Growth,* edited by Jay Shambaugh and Ryan Nunn, 71–121. Washington, DC: Brookings, The Hamilton Project.

Notowidigdo, Matthew J. Forthcoming. "The Incidence of Local Labor Demand Shocks." *Journal of Labor Economics.*

Oishi, Shigehiro. 2010. "The Psychology of Residential Mobility: Implications for the Self, Social Relationships, and Well-Being." *Perspectives on Psychological Science* 5 (1): 5–21.

Okun, Arthur. 1975. *Equity and Efficiency: The Big Tradeoff.* Washington, DC: The Brookings Institution.

Partridge, Mark D., and Dan S. Rickman. 2006. "An SVAR Model of Fluctuations in U.S. Migration Flows and State Labor Market Dynamics." *Southern Economic Journal* 72 (4): 958–80.

Paull, Evans. 2008. "The Environmental and Economic Impacts of Brownfields Redevelopment." Working draft for distribution, Northeast Midwest Institute, Washington, DC.

Persky, Joseph, Daniel Felsenstein, and Virginia Carlson. 2004. *Does "Trickle Down" Work? Economic Development and Job Chains in Local Labor Markets.* Kalamazoo, Michigan: The W.E. Upjohn Institute for Employment Research.

Pew Charitable Trusts. 2017. *How States Are Improving Tax Incentives for Jobs and Growth: A National Assessment of Evaluation Practices.* New York: The Pew Charitable Trusts.

Pierce, Justin R., and Peter K. Schott. 2017. "Trade Liberalization and Mortality: Evidence from U.S. Counties." Working paper, Board of Governors of the Federal Reserve System, Washington, DC.

Poterba, James M., and Lawrence H. Summers. 1995. "A CEO Survey of U.S. Companies' Time Horizons and Hurdle Rates." *MIT Sloan Management Review* 37 (1): 43.

Robey, Jim, Randall Eberts, Carlesa Beatty, Kathleen Bolter, Marie Holler, Brian Pittelko, Claudette Robey. 2018. "The National-Level Economic Impact of the Manufacturing Extension Partnership (MEP): Estimates for Fiscal Year 2017." Prepared for National Institute of Standards and Technology and Manufacturing Extension Partnership.

Rowthorn, Robert, and Andrew J. Glyn. 2006. "Convergence and Stability in U.S. Employment Rates." *Contributions in Macroeconomics* 6 (1): 1–43.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas and Matthew Sobek. 2020. IPUMS USA: Version 10.0 [American Community Survey, 2016 and 2018,]. Minneapolis, MN: IPUMS, 2020. https://doi.org/10.18128/D010.V10.0.

Saiz, Albert. 2010. "The Geographic Determinants of Housing Supply." *Quarterly Journal of Economics* 125 (3): 1253–96.

Schwandt, Hannes, and Till M. von Wachter. 2019. "Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets." *Journal of Labor Economics* 37 (S1): S161–S98.

Slattery, Cailin, and Owen Zidar. 2020. "Evaluating State and Local Business Incentives." *Journal of Economic Perspectives* 34 (2).

Smith, Adam. 1776 [1904]. *The Wealth of Nations*. Edited by Edwin Cannan, https://www.econlib.org/library/Smith/smWN.html.

Smith, Stephen C. 2018. "Development Economics Meets the Challenges of Lagging U.S. Areas: Applications to Education, Health and Nutrition, Behavior, and Infrastructure." In *Place-Based Policies for Shared Economic Growth*, edited by Jay Shambaugh and Ryan Nunn, 185–242. Washington, DC: Brookings Institution.

Story, Louise. 2012. "As Companies Seek Tax Deals, Governments Pay High Price". *The New York Times,* December 1, 2012.

Stuart, Bryan A. 2017. "Essays on the Economics of People and Places." Dissertation, University of Michigan Department of Economics.

Thomas, Kenneth. 2011. *Investment Incentives and the Global Competition for Capital*. London and New York: Palgrave Macmillan.

U.S. Bureau of Labor Statistics. 2015. "States: Employment status of the civilian noninstitutional population by sex, race, Hispanic or Latino ethnicity, and intermediate age, 2015 annual averages." United States Department of Labor. https://www.bls.gov/lau/ex14tables.htm (accessed May 27, 2020).

U.S. Bureau of Labor Statistics. 2018. "American Community Survey (ACS) Questions and Answers." *Local Area Unemployment Statistics*. https://www.bls.gov/lau/acsqa.html (accessed May 19, 2020).

U.S. Census Bureau. 2020. "2017 Census of Governments: Finance." https://www.census.gov/data/tables/2017/econ/gov-finances/summary-tables.html (accessed May 19, 2020).

U.S. Census Bureau. 2018. "2016 Update." *Business Dynamics Statistics (BDS)*. https://www.census.gov/data/tables/2016/econ/bds/2016-firm-and-estab-release-tables.html (accessed May 19, 2020).

U.S. Department of Treasury. 2016. "Tax Expenditures." Washington, DC: U.S. Department of the Treasury, Office of Tax Analysis. https://www.treasury.gov/resource-center/tax-policy/Documents/Tax-Expenditures-FY2018.pdf (accessed May 19, 2020).

Vance, J. D. 2016. *Hillbilly Elegy: A Memoir of a Family and Culture in Crisis*. New York: HarperCollins.

Wilson, William Julius. 1996. *When Work Disappears: The World of the New Urban Poor*. New York: Alfred A. Knopf.

Winnick, Louis. 1966. "Place Prosperity vs. People Prosperity: Welfare Considerations in the Geographic Redistribution of Economic Activity." In *Essays in Urban Land Economics*, Los Angeles: Real Estate Research Program, University of California.

Yagan, Danny. 2019. "Employment Hysteresis from the Great Recession." *Journal of Political Economy* 127 (5): 2505–58.

Zabek, Mike. 2019. "Local Ties in Spatial Equilibrium," Finance and Economics Discussion Paper 2019-080.

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 128–149

Place-Based Policies and Spatial Disparities across European Cities

Maximilian v. Ehrlich and Henry G. Overman

Spatial disparities in income and worklessness across areas of the European Union are profound and persistent. Concerns about these disparities and the appropriate policy response are longstanding. Two trends have re-energized popular and academic debate. One is economic: on some dimensions, disparities have stopped narrowing and started to grow. The other is political: some argue that persistent disparities cause discontent and help explain the rise in populist movements (Rodríguez-Pose 2018).

We focus on disparities in income and worklessness across EU metropolitan regions, commonly called "metros," using new definitions from OECD and Eurostat. As these metros account for around two-thirds of the population and for larger and growing shares of employment and GDP, their economic performance is crucial for understanding EU disparities. Focusing on them also narrows down the area-based policies that are relevant. It means we have less to say about rural-urban disparities which involve different economic mechanisms and policies.

Our metro definition is based on the so-called NUTS3 regions, which divide up Europe into areas of 150,000 to 800,000 people. Our data combines these areas into metro regions: groups of NUTS3 sharing a common labor market and meeting a minimum size threshold. We focus mostly on the "EU-15," which was the group of 15 countries in the EU at the end of 2003, before the EU expanded to central and eastern Europe. We also offer some comparisons to the "EU-28," referring to the

■ *Maximilian v. Ehrlich is Professor of Economics, University of Bern, Bern, Switzerland. Henry G. Overman is Professor of Economic Geography, London School of Economics, London, United Kingdom. Their email addresses are maximilian.vonehrlich@vwi.unibe.ch and h.g.overman@lse.ac.uk.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.128.

total number of EU countries before the departure of the United Kingdom, as well as some comparisons to the US economy.

We begin by providing evidence that differences in GDP per capita across EU-15 metros converged in the 1980s, stabilized in the 1990s and early 2000s, and have been diverging since the mid-2000s. We also show diverging patterns of worklessness.

We then turn to research in urban economics for theories and empirical evidence that help explain the factors driving these disparities. We will show that bigger cities pay higher wages (the "urban wage premium") because they make workers more productive. They also tend to attract more educated workers who are more productive and earn more. As a result, GDP per capita is higher in bigger cities. These two factors reinforce one another because the urban wage premium increases with education. Both factors play a role across EU metro areas in explaining the level and evolution of spatial disparities. We provide evidence that real estate costs increase with city size, with implications for real wage inequalities and whether area-level improvements in productivity capitalize into higher house prices. We also explore low mobility rates in Europe and differences in labor market regulations, which help explain why employment disparities are more pronounced than for income.

Do these profound spatial disparities justify place-based policies aimed at reducing them (Austin, Glaeser, and Summers 2018)? Neumark and Simpson (2015) provide a useful overview of the literature on place-based policies. We focus on several policies that target spatial differences directly. Our emphasis is on policies that work at broad spatial scales. We argue that it is important to differentiate between policies as they operate via different mechanisms and yield different trade-offs between spatial inequality and aggregate efficiency.

We start with EU cohesion policy. These convergence transfers appear to have fostered growth in supported areas and thus reduced income disparities, but the effects vary considerably across areas with the positive effects driven by areas with high human capital and high-quality local government. The evidence also finds decreasing returns from transfers. The changes in disparities over time suggest that the economic forces swamp the impact of EU policy. We then consider two major items of expenditure within total cohesion policy spending: transport and support for firms from capital subsidies. Finally, we consider enterprise zones and local employment multipliers for different kinds of private and public sector employment.

Europe has a long tradition of using place-based policies to support lagging regions and to address local downward spirals following structural change. While place-based policies did not prevent rising disparities in Europe, they may have modestly mitigated the increase.

The Evolution of Spatial Disparities across European Cities

A comprehensive literature discusses regional disparities in Europe. Much of this uses data on "NUTS2 regions" of 800,000 to 3 million inhabitants which also

determine eligibility for the main EU structural funds. In contrast, we use data on metro regions. As argued above, one reason for this is the economic importance of these metros, and their role in driving EU spatial disparities.

The other reasons for using metros are analytical, but important. The economic literature on spatial disparities emphasizes the need to think about the appropriate spatial unit. For example, functional urban areas tied together by flows of people and goods should be used to think about local labor markets. But, for many EU countries, NUTS2 regions do not approximate functional urban areas. For example, London is split into five NUTS2 regions and merging just these regions—so that the London metro is a single geographic unit—changes one commonly used measure of dispersion across the EU-15 by 29 percent. Moreover, NUTS2 cover disparate areas: comparing London, Paris, and Munich, with the agricultural areas of Ireland, the beaches of Andalusia, and the mountains of Tyrol. The economic theories that explain disparities across cities, countryside, beaches, and mountains would need to be quite broad. Such breadth also widens the relevant place-based policies.

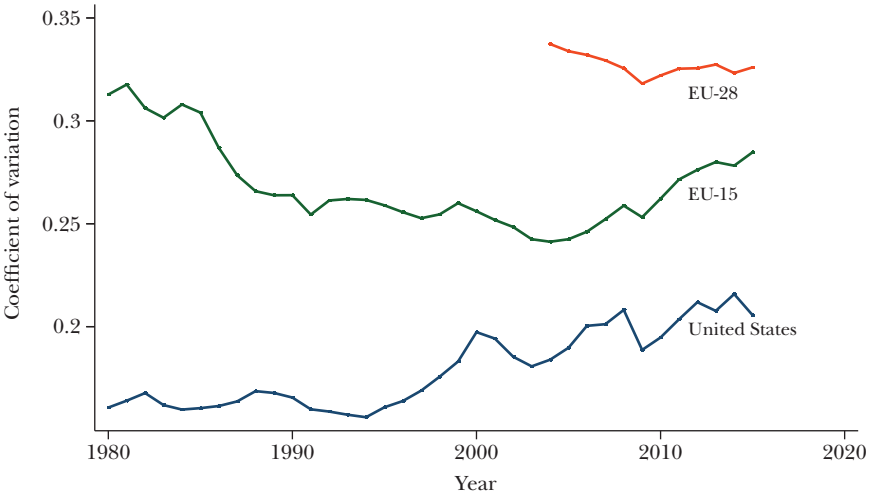
For these economic and analytical reasons, we focus on spatial disparities across metropolitan regions ("metros") using the recent EC/OECD specification (OECD 2019).[1] As described in the introduction, our data defines metros using NUTS3, or aggregates of NUTS3. For the EU-15 in 2015 (the latest date for which there is data), there are 226 metros with a minimum population of 250,000 and a maximum of 13.9 million. For the broader EU-28, we have 279 metros. In 2015, metros account for 64 percent of the population in the EU-15 (60 percent for the EU-28) and a higher share of employment and GDP.

One important headline indicator of disparities—because it determines eligibility for the main EU cohesion policy funds (discussed in detail later)—is whether a NUTS2 region has GDP per capita less than 75 percent of the EU average. Applying this indicator to EU-15 metros, 32 of 226 metros—home to 12.5 percent of the metro population—are below 75 percent of the average GDP per capita. For the EU-28, the corresponding figures are 51 out of 279 metros and 14 percent. In the United States, a similar proportion of metro areas (70 out of 384 as defined by the US Bureau of Economic Analysis) have per capita GDP that is 75 percent or less of the national average but account for only 7 percent of the metro population. In the EU, people are much more likely to live in poorer metros than in the United States. This hints at the role mobility plays in understanding EU disparities.

The coefficient of variation—the standard deviation divided by the mean—is a common measure of dispersion. Figure 1 plots the (unweighted) coefficients of variation of GDP per capita across EU-15, EU-28, and US metros over the last four decades. In 2015, the coefficient of variation was 0.28 for the EU-15 and 0.33 for the EU-28. EU disparities appear to be higher than their US equivalents, although the coefficients of variation are not directly comparable: for the United States, we

[1] The online Appendix available with this article at the *Journal of Economic Perspectives* website provides information on data sources, descriptive statistics and additional figures. It also provides a more detailed discussion of disparities across NUTS2 regions.

Figure 1
Coefficient of Variation of GDP Per Capita: EU-15, EU-28, and US Metros



Source: Based on authors' calculations.
Note: Calculations based on Eurostat and BEA data and metro definitions as described in the text. EU-15 and EU-28 calculations use GDP per capita; US uses income per capita.

used income (not GDP) per capita and Bureau of Economic Analysis (BEA) metros, rather than the OECD definition.[2] These differences are bigger if we include non-metro areas because the least productive rural areas in the EU are less productive (relative to the EU mean) than the least productive rural areas in the United States (relative to the US mean).

Variation across EU-15 and EU-28 countries explains around half the coefficient of variation for metro areas—44 percent and 50 percent, respectively (based on decomposing the squared coefficient of variation). EU-15 disparities fell in the 1980s, stabilized in the 1990s, fell again in the early 2000s, then increased from the mid-2000s and markedly after Europe's double-dip recession. For the EU-28, the coefficient of variation fell somewhat when new members joined and then remained at similar levels until 2015.

Disparities in income per capita across US metros started widening around 1995, roughly a decade before the EU-15. But since about 2004, the trends are relatively similar. From their lowest value in 2004, EU-15 disparities have increased by 18 percent, compared with 12 percent in the United States over the same period).

[2]We experimented with using data from the US Bureau of Economic Analysis, weighted by area shares, to approximate the OECD metro definition. However, the approximation is imprecise, so we focus on comparing trends rather than levels. The online Appendix provides a figure using comparable OECD metro area definitions applied to the United States (for a shorter time period), which confirms that the coefficient of variation for the EU-15 metros is 15 percent larger than for the United States (see Figure A1).

Figure 2
Coefficient of Variation of Worklessness: EU Metros

Source: Based on authors' calculations.
Note: Metro definitions as defined in the text.

For the EU-28, we observe a much higher level of disparity, but the short time series makes it hard to assess the longer run trend, which is why our focus is on the EU-15.

This rise in inequality across metros is especially striking because it follows a longer period of convergence across European regions in per capita income. Rosés and Wolf (2019) provide estimates of regional GDP per capita for a mixture of NUTS1 and NUTS2 regions (excluding Greece) and show a 31 percent decrease in the coefficient of variation between 1950 and 1980.

Another measure of convergence focuses on whether on average poor metros grow faster than rich metros by regressing growth rates of GDP per capita on initial levels, where the regression coefficient measures the extent to which regions are moving toward the mean level of per capita income (often referred to as beta-convergence). Running such regressions for 1980–2015 or for 1990–2015, we find evidence of significant mean-reversion, but for 2005–2015, we find divergence instead (see Figure A2 available in the online Appendix). Such findings reinforce the message that a longer-term pattern of mean-reversion of per capita income across the EU-15 has stalled and even reversed itself. This is similar to results for the United States (Ganong and Shoag 2017), although mean-reversion ended there around 15 years before it did in the European Union.

Other measures of economic performance show similar patterns. The rates of employment and worklessness (that is, of not working in the working-age population) also vary substantially. As shown in Figure 2, the coefficient of variation of worklessness for EU-15 metros increased from 0.31 in 2000 to 0.41 in 2015. The

level and trend are similar for the EU-28.[3] This variation in worklessness has been of long-standing interest in Europe and is receiving increased attention in the United States. For example, Austin, Glaeser, and Summers (2018) show that US disparities in worklessness rates are pronounced and have increased in the last decade.

Disparities in EU worklessness rates are more pronounced than those for GDP per capita: the coefficient of variation for per-capita GDP in 2015 is 0.28 and for worklessness is 0.41. As with GDP per capita, variation in per country worklessness explains around half the total variation (51 percent).

What Causes Geographical Disparities in Europe?

EU metros exhibit wide and persistent disparities in GDP per capita and in worklessness, and these disparities appear to be widening. To understand these disparities, the standard approach in urban economics is to think about firms and workers trading off productivity advantages of different cities for the costs of locating in those cities. (Urban amenities may play a role, too, but we sidestep that issue.)

Metro Disparities in Productivity and Land Prices

A substantial literature suggests urban size is an important source both of productivity advantages and of higher congestion and land costs. As an illustration, Figure 3 shows that city size is positively associated with GDP per worker and real estate prices. For 2015, regressing the log of GDP per worker on the log of city size gives an elasticity—the slope of the line in the figure—of 0.077. For the real estate index in 2011, the elasticity is 0.930.[4]

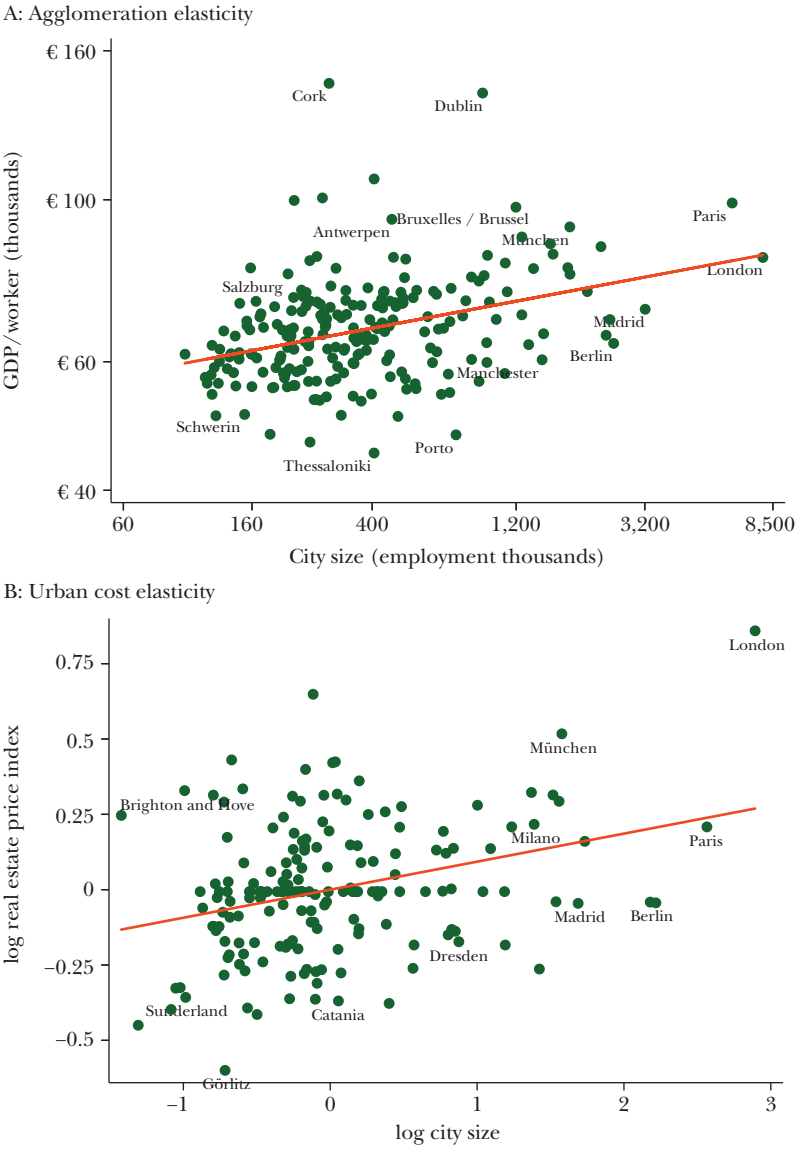
Because of the considerable wage premium earned by the "college educated," the relationship of GDP per capita with city size overstates productivity benefits if workers sort across cities in such a way that the higher-educated live in bigger more productive cities (as argued in Combes, Duranton, and Gobillon 2008). We see such sorting in our data: regressing metro GDP per worker on the share of population with tertiary education and absorbing country fixed effects gives a coefficient of 0.015 (that is, a 1 percentage point increase in the educated share increases GDP per worker by 1.5 percent).[5] Individual country-level studies control for such sorting on both observed characteristics (like the share of college educated) and

[3]For the EU-28, there is a longer time series of data on worklessness than there was for GDP per capita so we can look at the evolution over the same time-period as for the EU-15. Regressing the rate of worklessness in 2015 on the rate in 2005 gives a slope of 1.19 for EU-15 metros, suggesting that, as for GDP per capita, the recent past has seen divergence of worklessness. The same regression gives a coefficient of 1.07 for the EU-28.

[4]This second elasticity looks low compared to country-level estimates reported in Ahlfeldt and Pietrostefani (2019). This is not surprising, given that we pool together quite different data.

[5]Tertiary education data is only available from 2000 onwards for NUTS2. We compute the shares for metros by assigning each NUTS3 the corresponding NUTS2 education shares. For 14 metros, which only have data from 2005 on, we impute shares using a model with metro fixed effects and a linear time trend.

Figure 3
Agglomeration and Urban Costs: EU Metros



Source: Authors' calculations based on Boelmann and Schaffner (2019), Hilber and Mense (2020), Boeri et al. (2019) and other data sources detailed in the appendix.
Note: City size is number of workers in Panel A and population in Panel B. For Panel A, given variations in worklessness, we use GDP per worker and number of workers, rather than GDP per capita and population. Panel B uses data for France, Germany, Italy, Spain, and the UK and includes country fixed effects to account for differences in real estate price indices. Deviations in the log real estate index from the country mean are on the y-axis, deviations in log populations from the country mean are on the x-axis. Panel A uses data from 2015, Panel B from 2011 (Italy has no 2015 data). Results are robust to using 2015 and excluding Italy. For details, see the online Appendix.

Table 1
Agglomeration Elasticity: EU-15 Metros

Year	*Agglomeration elasticity*	*Agglomeration elasticity conditional tertiary education share*
1980	0.0429 (0.0260)	—
1990	0.0517 (0.0175)	—
2000	0.0778 (0.0136)	0.0764 (0.0135)
2010	0.0835 (0.0122)	0.0791 (0.0123)
2015	0.0774 (0.0132)	0.0686 (0.0134)

Source: Based on authors' calculations.
Note: Coefficients from regression of log GDP per worker on log number of workers controlling for share tertiary educated (column 2). Standard errors in parentheses.

unobserved characteristics (like the share with high ability) using individual panel data—that is, following specific workers over time. Unfortunately, no such panel data is available for the EU-15. However, if we re-estimate the relationship between GDP per capita and city size controlling for the share tertiary educated, the elasticity falls from 0.077 to 0.069.

Sorting and city size reinforce one another because more educated people live in more productive cities. Using US data, Moretti (2013) shows that the college wage premium is larger in big cities, a result we can replicate using less detailed individual level data from the EU. [6] The assortative matching of firms and workers may partially explain this effect (for discussion, see Card, Heining, and Kline 2013; Dauth et al. 2018).

Explaining the Changes in Disparities over Time?

If variations in city size and in the composition of educated workers help explain disparities across EU-15 metros, can a simple urban model also explain the changes over time?

Table 1 suggests a partial answer by looking at how the estimated elasticity of GDP per worker changes over time with respect to metro size. As convergence slowed and then reversed, the size elasticity increased markedly. In column 2, we control for sorting using the share of population with a tertiary education in periods when we have data. This has a relatively small effect on the agglomeration elasticities, although the effect does seem to be increasing over time. It is difficult to be precise because of the measurement error introduced by the way we must calculate tertiary education shares (see footnote 5).

[6] Using data from the EU Statistics on Income and Living Conditions (EU SILC), we run Mincer-style regressions including a city residence indicator interacted with a tertiary education indicator. The positive coefficient on the interaction suggests a higher tertiary education premium in cities, as shown in the online Appendix available at the *Journal of Economic Perspectives* website (see Table A3). Grujovic (2019) provides similar evidence with German data. Regressions using the EU SILC data show the high-skilled are 9.5 percent more likely to live in a city than the average and the effect has been increasing somewhat since the start of the data in 2005 (see Table A4).

We can look more directly at sorting by considering changes in the "college-educated" wage premium and in the spatial concentration of skilled workers. For some EU countries, the university graduate premium has increased (Machin and van Reenen 2007; Dustmann, Ludsteck, and Schönberg 2009) which directly increases disparities between smaller and bigger cities as the latter employ more highly educated workers.

Changes in the spatial concentration of highly educated workers reinforce the increase in the "college educated" wage premium. In EU-15 metros, the share of population with a tertiary education increased by about 10 percentage points between 2000 and 2015. This increase was not equally distributed across metros. Regressing the log growth of tertiary education shares on the log of initial population and including country fixed effects shows that a 10 percent increase in initial metro population is associated with a rise of 13.6 percent in the share tertiary educated over the period. That is, we see increased sorting of the more educated population consistent with US evidence (Moretti 2004; Berry and Glaeser, 2005). This increasing concentration of more educated workers is reflected in increased concentration of skill-intensive employment. For example, using patents as a proxy for skill-intensive employment, we see increased spatial concentration between the early 1990s and early 2010s.[7]

What explains the increasing concentration of more educated workers in big cities? One factor is the shift from manufacturing to knowledge-intensive services: the employment share of knowledge intensive services and high technology manufacturing increased in the EU from 2000 to 2015 by around 16 percent. This shift was caused by a mixture of increased globalisation (like the "China shock," as in Autor, Dorn, and Hanson 2013; Dauth and Südekum 2016) and technological change and increased automation (Acemoglu and Restrepo 2019; Dauth et al. 2019). As knowledge-intensive services employ more educated workers and benefit from higher agglomeration economies, this structural shift should see increased concentration of more educated workers in big cities.

An inelastic supply of housing in growing and more productive metros also plays a role. High house prices prevent the poor, who spend a higher income share on housing, from moving to more productive areas (Ganong and Shoag 2017). In some EU metros, land use constraints are highly restrictive and increase house prices (Hilber and Vermeulen 2016). For the EU countries in our data, real estate price increases are particularly pronounced in places with high initial GDP per worker.[8] For the United States, Hsieh and Moretti (2019) estimate the aggregate

[7] For details of the regression of the log growth of tertiary education shares on the log of initial population, see Figure A3 in the online Appendix available with this paper at the *Journal of Economic Perspectives* website. For details of evidence on increased spatial concentration, using patents as a proxy for skill-intensive employment, see Figure A4.

[8] For data showing correlations between real estate price increase and EU metro areas with high initial GDP per worker, see Figure A5 in the online Appendix available with this paper at the *Journal of Economic Perspectives* website.

GDP costs of the spatial misallocation resulting from such land use constraints, but no estimates are available for the EU.

Spatial Disparities in Worklessness

As is well known, differences in labor market institutions play an important role in explaining country variation in worklessness (in this journal, Siebert 1997). These institutions may also help explain why spatial disparities in worklessness are more pronounced. For example, nationally set minimum wages could increase worklessness in poorer areas: evidence for Germany suggests this happens in some low wage areas (Ahlfeldt, Roth, and Seidel 2019). Even without binding minimum wages, centralized wage bargaining may be a driver of spatial disparities in worklessness as such schemes prevent the adjustment of wages to regional productivity differences. Comparing Italy and Germany, Boeri et al. (2019) argue that centralized wage bargaining in Italy translates similar spatial variations in productivity into much smaller variation in nominal wages but much bigger variations in worklessness. Our results confirm the important role of labor market institutions: regressing metro worklessness rates against GDP per worker, we find a negative coefficient which is more than twice as large for countries with more centralized wage bargaining.[9]

Mobility and Spatial Disparities

According to Molloy, Smith, and Wozniak (in this journal, 2011), mobility in 2005 was significantly higher in the United States than in the European Union, which contributed to higher EU disparities. But in contrast to the United States where mobility rates have been falling, the EU trend is less clear, and mobility may have been increasing (EU Commission 2018). Fischer and Pfaffermayr (2018) suggests that labor mobility plays a small role in reducing EU disparities in per-capita GDP. Unfortunately, this increased mobility took place against a background of increasing concentration of economic activity and sorting of the high skilled toward big cities. There is also some evidence that regional transfers may slow down the adjustment that occurs via mobility (Egger, Eggert, and Larch 2014; Jofre-Monseny 2014).

Place-based Policies

So far, we have considered factors that explain disparities across EU metros and why these areas have stopped converging and have started to diverge. The rest of the paper considers place-based policies. We consider policies that *explicitly* target

[9] Conditional on country fixed effects, the effect of log GDP per worker on non-employment rates is –0.21 in the group of countries with more flexible regional wage bargaining (Austria, Germany, Denmark, Netherlands, Sweden) and –0.57 in the group with less flexible, more centralized wage bargaining (Belgium, Finland, France, Italy, Portugal, Slovenia). Both coefficients are significant at the 1 percent level.

the spatial allocation of economic activity. We will not discuss general national-level policies like schools funding, employment training, and others that directly target outcomes like education that matter for spatial disparities but aren't necessarily designed to target the issue of divergence. We focus on what we know about the impact of these policies on specific economic outcomes such as employment and how this depends on the economic forces driving spatial disparities that we discussed above.

These forces also affect the equity and efficiency of place-based policies. In distributional terms, the effect of policy will be partly determined by the mobility of individuals living in the area targeted and the housing supply elasticity (Kline and Moretti 2014). For example, with relatively elastic supply of labor across metros, but an inelastic housing supply, local benefits of spatial transfers are realized by landlords as they become capitalized into land prices. Firm and household mobility also increases the risk that if policy induces significant local employment effects in targeted areas, these may come at the cost of employment losses elsewhere. Displacement from richer to poorer metro areas will presumably narrow disparities.

The effect on overall output depends on whether agglomeration economies in targeted areas outweigh potential losses in non-targeted areas. Shifting investments and jobs from prosperous, productive areas to lagging, less productive regions is also likely to generate aggregate efficiency costs. The effect of displacement on aggregate welfare depends on equity considerations and also how it affects congestion externalities: for example, if displacement from richer to poorer cities reduces both congestion and agglomeration externalities, the net effect might decrease productivity, but increase welfare (for example, Fajgelbaum and Gaubert 2020; Henkel, Seidel, and Suedekum 2018). It is unlikely that policymakers have enough information to account for this potential mixture of externalities (Kline and Moretti 2014).

EU Cohesion Policy

Reducing spatial disparities in income and worklessness is a long-standing EU objective. Interventions directly funded by the European Union include investments in transport infrastructure and in local public goods and services—a mix of firm subsidies and human capital investments including employment training. There are three main funds: the European Social Fund, the European Regional Development Fund, and the Cohesion Fund. Other smaller funds also partly target less developed regions.

The cohesion policy budget for 2014–2020 is €645 billion (for a detailed description, see https://cohesiondata.ec.europa.eu/). Total expenditure is around one-third of the EU budget, which is small relative to total government expenditure. That said, the impact of EU policy is greater than the budget total suggests because EU state aid rules also restrict policy in member states. The lion's share of the budget (60 percent) goes towards "less developed" regions, with GDP per capita less than 75 percent of the EU average. Investments in transport infrastructure, research and development, and business support are the main expenditure categories accounting for 45–50 percent of the budget.

Various arguments are used to justify EU cohesion policies. One approach takes equity arguments used to justify policies to reduce disparities within nation-states and extends these to an EU-wide policy. For example, if all EU citizens should be entitled to similar public goods, EU policy may be justified as helping to equalize fiscal capacity.

From an efficiency perspective, cohesion policy could lead to higher aggregate output if there are diminishing returns to public investment, so that investing in areas with lower levels of public investment will produce larger gains. Or the EU might play a federal role coordinating investments that exert cross-area externalities. Or EU transfers may mitigate externalities from fiscal competition among jurisdictions.

An alternative argument makes the case for cohesion policies as a tool for advancing European integration. For example, transfers may build acceptance of the EU in new member states. This may be important if integration generates economic growth at the center at the expense of peripheral regions (Puga 2002) or if wealthier areas can set higher taxes because firms' desire to locate there reduces tax competition (Brülhart, Jametti, and Schmidheiny 2012).

The effects of EU cohesion policies have been studied extensively. Clear eligibility criteria, strictly applied and largely unchanged since 1989, allow for a (quasi-) experimental situation in which NUTS2 regions with GDP per capita slightly below the 75 percent threshold receive substantial transfers and can be compared to regions slightly above the threshold that do not. Becker, Egger, Ehrlich (2010) use this threshold to identify the effect of transfers using a regression discontinuity design. On average, transfers appear to have been effective in fostering growth in recipients and thus reducing disparities (Becker, Egger, Ehrlich 2010; Mohl and Hagen 2010; Pellegrini et al. 2013; Giua 2017).

However, the effects vary considerably across areas depending on local conditions. The positive effects are driven by regions with high human capital, as measured by education of the workforce, and high-quality local government, as measured by survey data about public services (Becker, Egger, Ehrlich 2013). Transfers are ineffective elsewhere. One potential reason is that while member states agree on strategy and budgets, project selection is done by regional authorities. Lower-quality local governments may choose ineffective policy. Or worse, may be more susceptible to increased rent-seeking activities and white-collar crime (Accetturo, de Blasio, and Ricci 2014; de Angelis, de Blasio, and Rizzica 2018).

The empirical evidence also suggests decreasing returns from cohesion transfers. Becker, Egger, Ehrlich (2012) and Cerqua and Pellegrini (2018) estimate the effects of transfer intensity (defined as transfers relative to local GDP). Their results imply that the marginal treatment effect declines with higher intensity and becomes zero at some "maximum desirable treatment intensity." One explanation is that limits to institutional capacity mean that additional subsidies are used with increasing inefficiency. Alternatively, the returns to investment may decrease in a way consistent with a neoclassical aggregate production function so that even high-quality governments see decreasing returns. The literature does not discriminate between these two explanations.

Finally, a key question is whether transfers lead to temporary or permanent improvements. The evidence is inconclusive but raises doubts that effects are long-lived. For example, case studies of the Italian Abruzzi region and the UK's South Yorkshire region, which lost eligibility in 1996 and 2006 (respectively) suggest improvements were temporary (Barone, David, de Blasio 2016; Di Cataldo 2017). Becker, Egger, Ehrlich (2018) look at all areas which lost eligibility, finding on average a reversion to pre-transfer trajectories once funds are cut.

The findings raise several questions about ways to improve cohesion policy. For example, should the EU allow for a longer transition period when areas become ineligible for subsidies? Are transfers well-targeted at investments that improve long-run growth? Given the importance of human capital to the effectiveness of subsidies—both directly in labor markets and indirectly through improving local institutional quality—perhaps human capital should be a higher priority than, say, infrastructure? Similarly, given that effectiveness decreases as transfers increase, would it make sense to transfer some subsidies from regions with a higher ratio of subsidies to GDP to regions with a lower ratio?

All the existing empirical evidence is for regions rather than metros. Given the economic importance of metros, and the difference between urban and rural economies, more should be done to understand the differential impacts of cohesion policy. As metros are on average more highly educated, and human capital and GDP per capita matter for effectiveness, the efficiency of the funds may be increased by targeting metros that are relatively high skilled compared to surrounding regions. At the same time, the increased sorting of more educated workers means that declining areas, which are losing their more educated labor force, will also be less able to transform transfers into growth. This raises questions around place-based policies that target skilled labor, an issue to which we return below.

So far, we have focused on the overall effect of EU cohesion policy considering the effects of transfers consisting of a bundle of interventions. Blouri and Ehrlich (2020) find that there is significant variation across interventions in their effects. Thus, we next consider the impact of different policies, drawing on cross-EU studies and papers looking at national policies.

Transport Infrastructure

A substantial share of EU cohesion spending is on transport infrastructure: 18 percent in 2014–2020, down from 25 percent in 2007–2013. Nation-state infrastructure investment is many times larger. One way of thinking about infrastructure projects is as a public capital input that makes firms more productive (Aschauer 1989). This assumes decreasing returns to infrastructure investment, consistent with the findings for EU cohesion policy. More recent literature has emphasized the importance of thinking about the transport network. Changing the network affects firm access to goods, markets, and input factors, as well as worker access to jobs. As these determine the relative attractiveness of places, infrastructure may affect the location of firms and workers, shaping the spatial distribution of activity. For

an overview of theory and empirics on the impact of transport infrastructure, see Redding and Turner (2015).

Recent empirical evidence has looked at these effects using the impact of road investments. For example, looking at incremental changes in UK road infrastructure, Gibbons et al. (2019) find substantial positive effects on area employment and number of establishments. While employment gains are largely driven by firm entry, some firm-level analysis also finds productivity increases for incumbent firms. Holl (2016) provides such evidence for improved highway access in Spain, which also increased economic activity close to highways. These studies show sizable local effects but may not identify aggregate effects when improvements impact the entire network.

A central aim of the European Union is to increase integration by lowering transaction costs, thus potentially causing fundamental changes in economic geography. For example, the Trans-European Network is a key project that aims to improve integration. However, there are long-running debates about the spatial effects of infrastructure in the "New Economic Geography" research (Krugman 1991; Fujita, Krugman, and Venables 1999; Puga 2002; Baldwin et al. 2003) For example, the "two-way roads" problem points out that transport improves the access of firms in less-developed regions to core markets but also increases the access of core firms to less-developed regions. As a result, transport investments may increase or decrease industrial concentration. Overall, this literature suggests that the effect on spatial disparities depends on several factors: the reduction in trade costs, wage differences, congestion costs, and mobility.

Unfortunately, the two-region structure common in these earlier models proved hard to adapt to multi-region settings and complex transport networks. More recent spatial economic models eliminate the possibility of multiple-equilibrium but more easily incorporate realistic multi-region geography (Allen and Arkolakis 2014; Redding and Rossi-Hansberg 2017). Once fitted to real world data, such models can assess the relative contribution of location, market access and local (perhaps innate?) productivity differences in explaining spatial disparities. They can also quantify the effects of changes to transport networks on the spatial distribution of employment, income, and aggregate welfare while allowing for displacement.

Santamaria (2019) uses this approach to quantify the welfare effects of reshaping the West German highway network after World War II and finds that this generated large, persistent income gains. Allen and Arkolakis (2019) derive a framework to compute the welfare impact of local infrastructure improvements in the presence of agglomeration and congestion externalities. Even without relocation, the welfare effects spread over the network through changes in price indices. Blouri and Ehrlich (2020) use a similar model to consider the general equilibrium impact of EU infrastructure investments. Investments increase local productivity and this combined with reduced transport costs, generates significant aggregate welfare gains—but only a relatively small reduction in income disparities. The utility-maximizing distribution of investments suggests that funds should

be redistributed towards more central regions and some border regions. Unfortunately, this redistribution is predicted to increase spatial income inequality, once again highlighting the trade-off between aggregate efficiency and spatial disparities.[10]

Can transport infrastructure investments explain the recent divergence across metro areas? Initial investments in the Trans-European Networks may have mostly completed national networks, and the associated increase in public capital stock could have driven between-country convergence in the 1980s. However, if later investment did more to complete the cross-country network or were targeted more to core areas, the contribution to convergence would be reduced.

Again, much of the available evidence considers regions rather than metros. This leaves questions about place-based policy that have not been widely addressed. If reallocating transport expenditure towards more central regions maximizes aggregate efficiency, would this also hold true within regions? Transport investment may also interact with educational composition: for example, public transport in big cities may attract more educated workers, thus helping explain increased sorting. This has not been studied for Europe as a whole, but Fretz, Parchet, and Robert-Nicoud (2017) study the effects of the construction of the Swiss highway network, showing that improved access for municipalities led to a significant increase in their share of high-income households.

Capital Subsidies and Enterprise Zones

Governments offer subsidies to specific firms, particularly in disadvantaged areas. Such subsidies raise two major concerns: the "deadweight" problem that they finance activities that firms would have undertaken anyhow; and the "displacement" problem that if subsidies encourage new activity in targeted areas, this may come at the cost of activity elsewhere.

Research seeking to understanding the deadweight and displacement effects from EU policies struggles with a lack of detailed data and substantial identification challenges (for example, see Bachtrögler and Hammer 2018; Beņkovskis et al. 2019).

Country-level studies have made more progress because one (unintended) consequence of EU state aid rules is that they induce exogenous variation to identify the impact of place-based capital subsidies. Some studies suggest that subsidies, if well designed, can alter firm behavior (which is to say that not all the impacts are deadweight). For example, Criscuolo et al. (2019) look at the impact of the UK's Regional Selective Assistance scheme, which provided discretionary grants to manufacturing firms in disadvantaged areas. The rules governing area eligibility are determined by EU rules. Thus, changes in EU rules provide a source of exogenous variation for estimating the impact on employment, unemployment, and other firm

[10] Further welfare gains can be realized by supranational coordination of infrastructure—for example, if governments tend to ignore foreign consumers when deciding on investment in border regions (Felbermayr and Tarassov 2019).

outcomes. Subsidies have large and statistically significant effects: increasing area-level manufacturing employment and decreasing unemployment. These effects are driven by small firms. Similar strategies have been used for other place-based capital schemes including the GRW rules that set maximum levels for different incentives across regions of Germany (Brachert, Dettmann, and Titze 2019; Etzel, Siegloch, and Wehrhöfer 2020) and Law 488/1992 that governs incentives received by firms to invest in lagging areas in Italy (Bronzini and de Blasio 2006). The results are not always positive. Bronzini and de Blasio (2006) find evidence of substantial deadweight and displacement: subsidized firms bring forward investment projects and gains may come at the expense of non-subsidized firms.

Enterprise zones, in most incarnations, offer a broader set of subsidies (not just capital subsidies), some of which may offer indirect support to firms (like relaxation of planning regulations) but in a specific area often much smaller than a metro area. Most of the literature on enterprise zones comes from the United States(for a summary, see Neumark and Simpson 2015), but a small literature considers the effect of European schemes, particularly the French Franches Urbaines (for example, Briant, Lafourcade, and Schmutz 2013; Mayer, Mayneris, and Py 2017; Givord, Rathelot, and Sillard 2013; Gobillon, Magnac, and Selod 2012).

One difference that emerges is that US enterprise zones have larger impacts on area unemployment, which may reflect the fact that some US schemes impose "local hiring conditions," (usually that a certain percentage of workers must live locally) which are not used in Europe.

Another difference is that deadweight and displacement concerns are more pronounced for enterprise zones than for place-based capital subsidies operating at broader spatial scales. One explanation is that the latter are often selective. For example, to be eligible to receive UK Regional Selective Assistance, a firm must demonstrate that it does *not* predominantly serve local markets. Such a requirement may reduce displacement compared to enterprise zones that provide non-discretionary subsidies to all firms within the zone. Another explanation is that a firm relocating to an enterprise zone within the same metro can access the same local labor markets and do business with existing customers and suppliers. In the absence of a local hiring requirement, it can even employ the same workers. This creates large incentives to relocate within metros. In contrast, firms relocating to take advantage of other place-based capital subsidies may need to move to different local labor markets and face differential access to customers and suppliers.

We have little evidence on the efficient spatial allocation of these area-based initiatives. As one example, Gaubert (2018) studies the location choice of heterogeneous firms when offered firm subsidies to locate in different size cities. In the model (calibrated to the French ZFU programme for urban tax-free zones), firm subsidies in small, less productive cities led to displacement, which has negative effects on aggregate productivity. Transfers to large, productive cities increase aggregate productivity.

The effects of these policies on spatial disparities will be modest. If the findings for UK Regional Selective Assistance generalize, selective (capital) subsidies may

reduce disparities in worklessness, but not GDP per capita. For the scale at which enterprise zones operate, and given the findings on displacement, it is unlikely that these have much impact on metro disparities in the European Union.

Local Employment Multipliers

Firm-level subsidies aim to support employment at an individual firm or to attract new employers to an area. This should directly increase local employment, providing that subsidized employment does not displace existing jobs. This increased local employment may generate additional jobs by increasing productivity (as in Greenstone, Hornbeck, and Moretti 2010) or demand for locally produced goods and services. These positive "multipliers" may be offset by general equilibrium effects that increase local wages or prices.

The literature on local multipliers assesses the net effect on local employment. The evidence considers multipliers from three kinds of employment: tradable sectors (that sell mostly outside the local economy); tradable skilled and high-tech sectors; and the public sector. The multiplier for jobs in tradable sectors on jobs in non-tradable sectors is the most frequently estimated. Estimates for Italy, Spain, Sweden, and the United Kingdom differ, although they are broadly in line with US estimates. This suggests that an additional tradable job creates between 0.5 and 1.5 extra jobs in the non-tradable sector. A smaller number of studies provide estimates for high-tech or high-skilled tradables, generally finding larger multipliers (again, consistent with US evidence).

The fact that these multipliers are higher might provide an additional justification, over and above the direct effect on innovation for policies that support the clustering and collaboration of firms in sectors that are intensive in research and development. However, evidence on the effectiveness of these policies is mixed. For example, for Germany, Falck, Heblich, and Kipar (2010) document positive effects on innovation, whereas Martin, Mayer, and Mayneris (2011) and Falck, Koenen, and Lohse (2019) tend to find no effects on regional employment in France and Germany, respectively. Moreover, these studies ignore the negative aggregate effect of spreading out activities that may benefit from large agglomeration economies. It also ignores the possibility that price effects, like higher prices of housing, may outweigh any employment effects for the lower skilled (Lee and Clarke 2019).

Decisions about public sector employment allow governments to affect the spatial allocation of employment directly. For example, central government employment is usually concentrated in the capital city. Reallocation of public sector employment from richer to poorer areas provides a direct mechanism for reducing disparities.

Some studies estimate multiplier effects for these public sector jobs (Faggio and Overman 2014). The What Works Centre for Local Economic Growth (2019) identified six such studies. Results are mixed, with two finding negative effects on private sector employment (that is, crowding out), one finding no effect, and three finding positive multipliers. Two of these three report crowding out for manufacturing, offset by a positive multiplier on services. Increases in wages or house prices seem to underpin these crowding out effects. Overall, estimated public sector

multipliers are smaller than private sector ones. One explanation is that public sector employers may have weaker input-output linkages with local firms. Another is that salaries are relatively high in relocated public sector jobs, consistent with both larger price effects on wages and housing and higher levels of crowding out.

Conclusion

Spatial disparities across EU metro areas are profound, persistent, and may be widening. Thinking about the role of metros and the sorting of workers helps us to better understand these disparities and the effect of different policies and complements the extensive literature on regional disparities. The findings that EU support is more effective in higher educated regions, on the intensity of transfers and the impact of transport, raise questions about whether funds should be targeted more at metros. Regardless of the intervention, our understanding of many place-based policies is improved if we think about the effects from a metro perspective.

Our discussion has raised several questions without answering them, and here is one more. At least as far back as Akerlof et al. (1991), economists have raised the possibility of employment subsidies to help address EU disparities and reduce the risk of "downward spirals" arising from large localized negative shocks. But the emphasis of EU cohesion policy has remained on infrastructure investment and physical capital subsidies. Perhaps the set of cohesion policy instruments needs to be expanded?

Historically, arguments between proponents of place-based or place-blind policies have been conducted as an either-or debate. In a world where some people are mobile, and others are not, we do not find this distinction helpful. Instead we need to understand the impacts of a range of different policies regardless of whether they are targeted at people or at places. The cost-effectiveness, the consequences for spatial disparities, and the benefits for different kinds of people living in different places are likely to vary significantly across policies. It is unlikely that a priori classifications of policy as place-based or place-blind will be very informative about these differential impacts on redistribution and aggregate efficiency or the tradeoffs between them.

■ *We thank the editors Gordon Hanson, Enrico Moretti, Heidi Williams, and Timothy Taylor for many very helpful comments. We benefited from comments by Gabriel Ahlfeldt, Guido de Blasio, Gilles Duranton, Tobias Seidel, Jens Suedkum, Paul Swinney, and Elisabet Viladecans-Marsal. We thank Christian Hilber, Stefan Fahrlaender and Johanna Posch for sharing data with us.*

References

Accetturo, Antonio, Guido de Blasio, and Lorenzo Ricci. 2014. "A Tale of an Unwanted Outcome: Transfers and Local Endowments of Trust and Cooperation." *Journal of Economic Behavior & Organization* 102: 74–89.

Acemoglu, Daron, and Pascual Restrepo. 2019. "Automation and New Tasks: How Technology Displaces and Reinstates Labor." *Journal of Economic Perspectives* 33 (2): 3–30.

Ahlfeldt, Gabriel, Fabian Bald, Duncan Roth, and Tobias Seidel. 2019. "The Spatial Equilibrium with Migration Costs." Unpublished. http://personal.lse.ac.uk/ahlfeldg/WP/GA_FB_DR_TS_-_SEMC.pdf.

Ahlfeldt, Gabriel, and Elisabetta Pietrostefani. 2019. "The Compact City in Empirical Research: A Quantitative Literature Review." Spatial Economic Research Centre Discussion Paper 215.

Ahlfeldt, Gabriel, Duncan Roth, and Tobias Seidel. 2019. "Employment-Maximizing Minimum Wages." Unpublished. http://personal.lse.ac.uk/ahlfeldg/WP/GA_DR_TS_-_MW.pdf.

Akerlof, George A., Andrew K. Rose, Janet, L. Yellen, and Helga Hessenius. 1991. "East Germany in from the Cold: The Economic Aftermath of Currency Union." *Brookings Papers on Economic Activity* 22 (1): 1–105.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *Quarterly Journal of Economics* 129 (3): 1085–140.

Allen, Treb, and Costas Arkolakis. 2019. "The Welfare Effects of Transportation Infrastructure Improvements." NBER Working Paper 25487.

Aschauer, David Alan. 1989. "Is Public Expenditure Productive?" *Journal of Monetary Economics* 23 (2): 177–200.

Austin, Benjamin, Edward Glaeser, and Lawrence Summers. 2018. "Saving the Heartland: Place-Based Policies in 21st Century America." *Brookings Papers on Economic Activity* 49 (1): 151–232.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States" *American Economic Revie*w 103 (6): 2121–68.

Bachtrögler, Julia, and Christoph Hammer. 2018. "Who are the Beneficiaries of the Structural Funds and the Cohesion Fund and How Does the Cohesion Policy Impact Firm-Level Performance?" OECD Economics Department Working Paper 1499.

Baldwin, Richard, Rikard Forslid, Philippe Martin, and Gianmarco Ottaviano. 2003. *Economic Geography and Public Policy.* Princeton, NJ: Princeton University Press.

Barone, Guglielmo, Francesco David, and Guido de Blasio. 2016. "Boulevard of Broken Dreams. The End of EU Funding (1997: Abruzzi, Italy)." *Regional Science and Urban Economics* 60: 31–8.

Becker, Sascha O., Peter H. Egger, and Maximilian v. Ehrlich. 2010. "Going NUTS: The Effect of EU Structural Funds on Regional Performance." *Journal of Public Economics* 94 (9): 578–90.

Becker, Sascha O., Peter H. Egger, and Maximilian v. Ehrlich. 2012. "Too Much of a Good Thing? On the Growth Effects of the EU's Regional Policy." *European Economic Review* 56 (4): 648–68.

Becker, Sascha O., Peter H. Egger, and Maximilian v. Ehrlich. 2013. "Absorptive Capacity and the Growth and Investment Effects of Regional Transfers: A Regression Discontinuity Design with Heterogeneous Treatment Effects." *American Economic Journal: Economic Policy* 5 (4): 29–77.

Becker, Sascha O., Peter H. Egger, and Maximilian v. Ehrlich. 2018. "Effects of EU Regional Policy: 1989–2013." *Regional Science and Urban Economics* 69: 143–52.

Berry, Christopher R., and Edward L. Glaeser. 2005. "The Divergence of Human Capital Levels across Cities." NBER Working Paper 11617.

Beņkovskis, Konstanīns, Oļegs Tkačevs, and Naomitsu Yashiro. 2019. "Importance of EU Regional Support Programmes for Firm Performance." *Economic Policy* 34 (98): 267–313.

Blouri, Yashar, and Maximilian v. Ehrlich .2020. "On the Optimal Design of Place-Based Policies: A Structural Evaluation of EU Regional Transfers." *Journal of International Economics* 125: Article 103319.

Boelmann, Barbara, and Sandra Schaffner. 2018. "FDT Data Description: Real-Estate data for Germany (RWI-GEO-RED). Advertisements on the Internet Platform Immobilienscout24." *Rheinisch-Westfälisches Institut für Wirtschaftsforschung, Essen Projektberichte.*

Boeri, Tito, Andrea Ichino, Enrico Moretti, and Johanna Posch. 2019. "Wage Equalization and Regional Misallocation: Evidence from Italian and German Provinces." NBER Working Paper 25612.

Brachert, Matthias, Eva Dettmann, and Mirko Titze. 2019. "The Regional Effects of a Place-Based Policy—Causal Evidence from Germany." *Regional Science and Urban Economics* 79: Article 103483.

Briant, Anthony, Miren Lafourcade, and Benoit Schmutz. 2013. "Can Tax Breaks Beat Geography? Lessons from the French Enterprise Zone Experience." *American Economic Journal: Economic Policy* 7 (2): 88–124.

Bronzini, Raffaello, and Guido de Blasio. 2006. "Evaluating the Impact of Investment Incentives: The Case of Italy's Law 488/1992." *Journal of Urban Economics* 60 (2): 327–49.

Brülhart, Marius, Mario Jametti, and Kurt Schmidheiny. 2012. "Do Agglomeration Economies Reduce the Sensitivity of Firm Location to Tax Differentials." *Economic Journal* 122 (563): 1069–93.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and The Rise of West German Wage Inequality." *Quarterly Journal of Economics* 128 (3): 967–1015.

Cerqua, Augusto, and Guido Pellegrini. 2018. "Are We Spending Too Much to Grow? The Case of Structural Funds." *Journal of Regional Science* 58 (3): 535–63.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics* 63 (2): 723–42.

Criscuolo, Chiara, Ralf Martin, Henry G. Overman, and John Van Reenen. 2019. "Some Causal Effects of an Industrial Policy." *American Economic Review* 109 (1): 48–85.

Dauth, Wolfgang, Sebastian Findeisen, Enrico Moretti, and Jens Südekum. 2018. "Matching in Cities." NBER Working Paper 25227.

Dauth, Wolfgang, Sebastian Findeisen, Jens Suedekum, and Nicole Woessner. 2019. "The Adjustment of Labor Markets to Robots." Unpublished. https://sfndsn.github.io/downloads/AdjustmentLaborRobots.pdf.

Dauth, Wolfgang, and Jens Südekum. 2016: "Globalization and Local Profiles of Economic Growth and Industrial Change." *Journal of Economic Geography* 16 (5): 1007–34.

De Angelis, Ilaria, Guido de Blasio, and Lucia Rizzica. 2018. "On the Unintended Effects of Public Transfers: Evidence from EU Funding to Southern Italy." Banca d'Italia Working Paper 1180.

Di Cataldo, Marco. 2017. "The Impact of EU Objective 1 Funds on Regional Development: Evidence from the U.K. and the Prospect of Brexit." *Journal of Regional Science* 57 (5): 814–39.

Dijkstra, Lewis, Hugo Poelman, and Paolo Veneri. 2019. "The EU-OECD Definition of a Functional Urban Area." OECD Regional Development Working Paper 2019/11.

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg. 2009. "Revisiting the German Wage Structure." *Quarterly Journal of Economics* 124 (2): 843–81.

Egger, Peter H., Wolfgang Eggert, and Mario Larch. 2014. "Structural Operations and Net Migration across European Union Member Countries." *Review of International Economics* 22 (2): 352–78.

Etzel, Tobias, Sebastian Siegloch, and Wehrhöfer, Nils. 2020. "Efficiency and Equity Effects of Place-Based Policies." Unpublished.

European Commission. 2018. *Annual Report on Labour Mobility.* Brussels, Belgium: European Commission.

Faggio, Giulia, and Henry G. Overman. 2014. "The Effect of Public Sector Employment on Local Labour Markets." *Journal of Urban Economics* 79: 91–107.

Fajgelbaum, Pablo D., and Cecile Gaubert. 2020. "Optimal Spatial Policies, Geography, and Sorting." *Quarterly Journal of Economics* 135 (2): 959–1036.

Falck, Oliver, Stephan Heblich, and Stefan Kipar. 2010. "Industrial Innovation: Direct Evidence from a Cluster-Oriented Policy." *Regional Science and Urban Economics* 40 (6): 574–82.

Falck, Oliver, Johannes Koenen, and Tobias Lohse. 2019. "Evaluating a Place-Based Innovation Policy: Evidence from the Innovative Regional Growth Cores Program in East Germany." *Regional Science and Urban Economics 79:* Article 103480.

Felbermayr, Gabriel J., and Alexander Tarasov. 2019. "Trade and the Spatial Distribution of Transport Infrastructure." CESifo Working Paper 5634.

Fischer, Lorenz Benedikt, and Michael Pfaffermayr. 2018. "The More the Merrier? Migration and Convergence among European Regions." *Regional Science and Urban Economics* 72: 103–14.

Fretz, Stephan, Raphaël Parchet, and Frédéric Robert-Nicoud. 2017. "Highways, Market Access, and Spatial Sorting." Spatial Economics Research Centre Discussion Paper 227.

Fujita, Masahisa, Paul Krugman, and Anthony Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade.* Cambridge, MA: MIT Press.

Ganong, Peter, and Daniel Shoag. 2017. "Why Has Regional Income Convergence in the US Declined?" *Journal of Urban Economics* 102: 76–90.

Gaubert, Cécile. 2018. "Firm Sorting and Agglomeration." *American Economic Review* 108 (11): 3117–53.

Gibbons, Stepehn, Teemu Lyytikäinen, Henry G. Overman, and Rosa Sanchis-Guarner. 2019. "New Road Infrastructure: The Effects on Firms." *Journal of Urban Economics* 110: 35–50.

Giua, Mara. 2017. "Spatial Discontinuity for the Impact Assessment of the EU Regional Policy: The Case of Italian Objective 1 Regions." *Journal of Regional Science* 57 (1): 109–31.

Givord, Pauline, Roland Rathelot, and Patrick Sillard. 2013. "Place-Based Tax Exemptions and Displacement Effects: An Evaluation of the 'Zones Franches Urbaines' Program." *Regional Science and Urban Economics* 43 (1): 151–63.

Gobillon, Laurent, Thierry Magnac, and Harris Selod. 2012. "Do Unemployed Workers Benefit from Enterprise Zones? The French Experience." *Journal of Public Economics* 96 (9–10): 881–92.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings." *Journal of Political Economy* 118 (3): 536–98.

Grujovic, Anja. 2019. "Tasks, Cities, and Urban Wage Premia." Centre pour la Recherche Economique et ses Applications Document de Travail 1807.

Henkel, Marcel, Tobias Seidel, and Jens Suedekum. 2018. "Fiscal Equalization in the Spatial Economy." CESifo Working Paper 7012.

Hilber, Christian A.L., and Andreas Mense. 2020. "Decomposing Local House Price and Rent Dynamics in England." Unpublished.

Hilber, Christian A.L., and Wouter Vermeulen. 2016. "The Impact of Supply Constraints on House Prices in England." *Economic Journal* 126 (591): 358–405.

Holl, Adelheid, 2016. "Highways and Productivity in Manufacturing Firms." *Journal of Urban Economics* 93: 131–51.

Hsieh, Chang-Tai, and Enrico Moretti. 2019. "Housing Constraints and Spatial Misallocation." *American Economic Journal: Macroeconomics* 11 (2): 1–39.

Jofre-Monseny, Jordi. 2014. "The Effects of Unemployment Protection on Migration in Lagging Regions." *Regional Science and Urban Economics* 83: 73–86.

Kline, Patrick, and Enrico Moretti. 2014. "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." *Annual Review of Economics* 6: 629–62.

Krugman, Paul. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy* 99 (3): 483–99.

Lee, Neil, and Stephen Clarke. 2019. "Do Low-Skilled Workers Gain from High-Tech Employment Growth? High-Technology Multipliers, Employment and Wages in Britain." *Research Policy* 48 (9): 103803.

Machin, Stephen, and John van Reenen. 2007. "Changes in Wage Inequality." Centre for Economic Performance Special Paper 18.

Martin, Philippe, Thierry Mayer. and Florian Mayneris. 2011. "Public Support to Clusters: A Firm Level Study of French 'Local Productive Systems'." *Regional Science and Urban Economics* 41 (2): 108–23.

Mayer, Thierry, Florian Mayneris, and Loriane Py. 2017. "The Impact of Urban Enterprise Zones on Establishment Location Decisions and Labor Market Outcomes: Evidence from France." Journal of Economic Geography 17 (4): 709–52.

Mohl, Philipp, and Tobias Hagen. 2010. "Do EU Structural Funds Promote Regional Growth? New Evidence from Various Panel Data Approaches." *Regional Science and Urban Economics* 40 (5): 353–65.

Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic Perspectives* 25 (3): 173–96.

Moretti, Enrico. 2004. "Human Capital Externalities in Cities." In *Handbook of Urban and Regional Economics*, Vol. 4, edited by J. Vernon Henderson and Jacques-Francois Thisse, 2243–91. Amsterdam: North Holland.

Moretti, Enrico. 2013. "Real Wage Inequality." *American Economic Journal: Applied Economics* 5 (1): 65–103.

Neumark, David, and Helen Simpson. 2015. "Place-Based Policies." In *Handbook of Regional and Urban Economics*, Vol. 5B, edited by Gilles Duranton, Vernon Henderson, and William Strange, 1197–287. Amsterdam: Elsevier.

Pellegrini, Guido, Flavia Terribile, Ornella Tarola, Teo Muccigrosso, and Federica Busillo. 2013. "Measuring the Effects of European Regional Policy on Economic Growth: A Regression Discontinuity Approach." *Papers in Regional Science* 92 (1): 217–33.

Puga, Diego. 2002. "European Regional Policies in Light of Recent Location Theories." *Journal of Economic Geography* 2 (4): 373–406.

Redding, Stephen J., and Esteban Rossi-Hansberg. 2017. "Quantitative Spatial Economics." *Annual Review of Economics* 9: 21–58.

Redding, Stephen J., and Matthew A. Turner. 2015. "Transportation Costs and the Spatial Organization of

Economic Activity." In *Handbook of Regional and Urban Economics*, Vol. 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 1339–98. Amsterdam: Elsevier.

Rodríguez-Pose, Andrés. 2018. "The Revenge of the Places That Don't Matter (and What to Do About It)." *Cambridge Journal of Regions, Economy and Society* 11 (1): 189–209.

Rosés, Joan Ramón, and Nikolaus Wolf, eds. 2019. *The Economic Development of Europe's Regions. A Quantitative History since 1900*. Abingdon, UK: Routledge.

Santamaria, Marta. 2019. "The Gains from Reshaping Infrastructure: Evidence from the Division of Germany." Unpublished. https://drive.google.com/file/d/1lxynf4z09jtWFuJhxHD96pm7ssikUZBE.

Siebert, Horst 1997. "Labor Market Rigidities: At the Root of Unemployment in Europe." *Journal of Economic Perspectives* 11 (3): 37–54.

What Works Centre for Local Economic Growth. 2019. *Local Multipliers*. London: What Works Centre for Local Economic Growth.

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 150–173

Urbanization in the Developing World: Too Early or Too Slow?

J. Vernon Henderson and Matthew A. Turner

Most regions of the world seem fully urbanized. North America, Europe, Latin America and the Caribbean, and West Asia all have shares of the population living in urban areas over 68 percent, with most regions near 80 percent. They also have small annual growth rates in this share, all under 0.62 percent a year and most near 0.25 percent (United Nations 2018). East Asia still has rapid urbanization, but its population is now over 60 percent urbanized and should soon top 70 percent, as in more developed regions. North Africa is only just over 50 percent urban, but that number is stable with little further urbanization. The global frontier of rising urbanization is sub-Saharan Africa (urbanization rate of 40 percent, annual growth rate of 1.4 percent), South Asia (urbanization rate of 36 percent, annual growth rate of 1.2 percent) and South-East Asia (urbanization rate of 49 percent, annual growth rate of 1.3 percent). Urbanization in these regions, and in sub-Saharan Africa in particular, will be the focus of much of our attention.

■ *J. Vernon Henderson is School Professor of Economic Geography, London School of Economics, London, United Kingdom. Henderson is also a Research Fellow of the Centre for Economic Policy Research. Matthew A. Turner is Professor of Economics, Brown University, Providence, Rhode Island. Turner is also Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts, and Senior Research Fellow, Property and Environment Research Center, Bozeman, Montana. Both authors are Research Affiliates, International Growth Center, London, United Kingdom. Their email addresses are J.V.Henderson@lse.ac.uk and matthew_turner@brown.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.150.

To understand the forces driving urbanization in developing countries, we begin by documenting key patterns and puzzles about urbanization and population density that have emerged in the literature. The classic economic model of urbanization is a story of technological change and structural transformation. Over generational time scales, people move from rural farms to urban factories in response to higher productivity in cities (for a review of this literature, see Desmet and Henderson 2015). East Asia and regions that urbanized prior to the late twentieth century seem to follow this path. However, sub-Saharan Africa is different. There, many countries are urbanizing "early"—that is, urbanizing at levels of per capita income generally far lower than when previous regions urbanized. Moreover, many cities in sub-Saharan Africa are growing without the expected increase in manufacturing or decline in agriculture. Perhaps related, many more farmers are living in urban areas in sub-Saharan Africa than we would predict from observing cities in other places and times. For a review on sub-Saharan Africa urbanization, see Henderson and Kriticos (2018).

Next, we offer evidence about costs and benefits of living in urban and rural areas in developing countries. In the first part of the paper, we rely on the dual sector model and its division into urban and rural. However, in the presentation of facts about costs and benefits, as in models in the modern literature (for example, Michaels, Rauch, and Redding 2012), we treat space as a continuum. We explore how various outcomes change with population density estimated at the level of a global grid of one-kilometer squares, using the Global Human Settlements Layer (GHSL) data. We show that significant fractions of the urban population in developing countries live at high densities that are practically nonexistent in the developed world. This suggests that the impacts of such densities can only be studied by looking at the developing world.

For data on a variety of outcomes, we use three geocoded surveys: the World Bank Living Standards and Measurement Survey (LSMS) for data on income and wages; the Demographic and Health Surveys (DHS) for data on a variety of amenities; and the Afrobarometer surveys for data on crime. We find that the high densities of developing world cities, in Africa and South Asia in particular, are associated with many benefits, including higher incomes and access to electricity, clean water, and inoculations. However, they also entail costs, including higher incidence of lifestyle diseases, poorer child health outcomes, and greater exposure to crime.

Finally, we turn to a discussion of spatial equilibrium and the differences between rural and urban life that inform the rational choice of location. The classic Roback (1982) model of spatial equilibrium suggests that people will move between rural and urban locations until they have equalized utility of living in the two areas. However, our results tend to confirm earlier findings that incomes and wages are far higher in urban than in rural areas of developing countries (for earlier work, useful starting points are Gollin, Lagakos, and Waugh 2013; Chauvin et al. 2017). Like earlier work, we also find that many urban amenities in rapidly urbanizing countries dominate those in rural areas (for example, see Gollin, Kirchberger, and Lagakos 2017). In short, utility levels seem higher in urban areas of developing countries.

To account for this, the classic model of spatial equilibrium has been modified in various ways. For example, structural modelling now incorporates moving costs or affinities for particular locations (for example, Tombe and Zhu 2019; Balboni 2019; Bryan and Morten 2019). Perhaps people are so attached to rural locations, or the rural-to-urban move is so costly, that the large apparent benefits of urban life are still not large enough to motivate migrating. It may also be that certain negative aspects of urban living may play a larger role in people's decision-making than previously recognized.

Much remains to be understood about how the drivers of urbanization in sub-Saharan Africa and other developing countries differ from the classic model of rural migrants heading for urban manufacturing jobs. But looking at the urban/rural gaps in income and amenities, we ask whether the true puzzle is not whether urbanization is happening too early, but rather, why it is not happening even faster.

Some Distinctive Patterns of Urbanization in Developing Countries

Early Urbanization?

Many low-income countries today are urbanizing "early"—that is, at historically low levels of income—with the prime examples being countries in sub-Saharan Africa (Lall, Henderson, and Venables 2017; Bryan, Glaeser, and Tsivanidis 2019). The nations of sub-Saharan Africa surpassed 40 percent urban share in 2010 at a GDP per capita of $1,481. For comparison, Latin America passed the 40 percent mark in 1950 at a GDP per capita of $2,500, while East Asia surpassed a 40 percent urban share in 2000 at a per capita GDP of $5,451.[1] For reference, in 1900, per capita in Western Europe was at least twice that of sub-Saharan Africa today.

Why might urbanizing while poor matter? Early urbanization poses enormous challenges in governance. Poor countries cannot afford the ideal investments required to deal with the negative externalities of dense cities and are always playing a game of catch-up with rapid industrialization. Clustering of employment requires expensive transportation infrastructure to allow large numbers of workers to reach firms in city centers or peripheral industrial and commercial zones as well as to allow firms to get their goods to markets (Fujita and Ogawa 1982; Heblich, Redding, and Sturm 2018; Akbar et al. 2018; Tsivanidis 2019). The sewer systems and safe water supplies required to improve health and reduce mortality from disease (Kappner 2019) at high population densities are also expensive.

The problem goes beyond a simple lack of funds. Cities require institutions to collect taxes, keep order, and govern land. It is natural to suspect that the institutions

[1] All GDP numbers here are expressed in 1990 dollars at the purchasing power parity exchange rate, based on Bolt, Timmer, and van Zanden (2014). In 2010, the comparable number for South and South East Asia was $3,537, with South Asia still well under 40 percent urbanized today and South East Asia only having passed that mark in about 2005.

Table 1
Share of Manufacturing in GDP by Region and Year

Region	*1990*	*2000*	*2010*	*2017*
East Asia	24.6	25.2	27.6	27.4
South East Asia	22	24.8	22.6	20.9
Latin America and Caribbean	20.7	17.9	15.7	15.2
North Africa	17.6	17.9	16	16
Europe	17.5	15.3	11.9	11.8
South Asia	15.9	15.6	16.1	14.4
West Asia	14.4	13.2	12.1	13.8
Sub-Saharan Africa	13.8	11.6	8	9

Note: Data from the World Development Indicators 2018 are organized by UN regions. The table reports regional weighted averages using weights based on country share of regional GDP in 2017. Data cover 126 countries in a consistent sample over time. The Middle East is part of West Asia (not North Africa). Oceania is excluded.

and state capacity in these newly urbanizing areas reflect the lower income and education of the population.

What Is Driving Developing Country Urbanization?

The classic dual sector model of urbanization predicts that urban populations arise as farmers move to cities to work in factories making manufactured goods. Countries like Brazil and Argentina each had about 30 percent of GDP in manufacturing in 1980 even as urbanization was starting to slow, while China was over 40 percent in 1979 as urbanization was just taking off (World Bank 1981). Table 1 shows regional patterns for 1990 onwards where we have a large enough sample of countries reporting data for all listed years. As of 2017, East Asia has 27 percent of GDP from manufacturing, China about 29 percent, and South East Asia 21 percent. East and South East Asia maintained high manufacturing shares over the whole 1990 to 2017 time period. Latin America's manufacturing share started at over 20 percent in 1990 and declined to just over 15 percent.

In contrast, the 33 countries of sub-Saharan Africa that our data describe (excluding South Africa) have the lowest regional share of manufacturing worldwide in 1990—a share that has only declined over time. While other regions have experienced declines in manufacturing share, they tend to be countries with high income levels that are deindustrializing in favour of traded services. In general, most of sub-Saharan Africa has never developed a manufacturing sector beyond production of traditional goods for within-country consumption.

In short, sub-Saharan Africa and parts of South Asia have relatively few manufacturing employees and their numbers are growing slowly. So what is driving urbanization in these regions? We consider several possibilities, but there is no agreed-upon answer.

One possibility is that the current wave of developing country urbanization is being led by consumption opportunities, including urban amenities rather than production. A literature on "consumer cities" began with Glaeser, Kolko, and Saiz

(2001) and was extended to developing countries by Gollin, Jedwab, and Vollrath (2016). The latter paper demonstrates that in Asia and Latin America, there is a strong positive correlation between urban share and the GDP share of manufacturing and services. However, no such correlation exists in Africa and the Middle East. They conclude that urbanization in sub-Saharan Africa and the Middle East is driven by rents from natural resource exports, which they conjecture are distributed in cities. Such rents then can fund civil servants and urban services as well generate demand for urban private services. Also, one may draw a connection between natural resources rents and low manufacturing, based on the argument that revenues from natural resource exports affect exchange rates and wage costs, crowding out manufacturing and its technological spillover benefits (Sachs and Warner 2001; Ismail 2010; Alcott and Keniston 2017).

Henderson and Kriticos (2018) reexamine the consumer city argument. While they confirm the finding that urbanization in sub-Saharan Africa is not correlated with the manufacturing and services share of GDP, they find that increases in natural resource rents are also not associated with increased urbanization in Latin America and sub-Saharan Africa. More generally, countries without natural resource rents are urbanizing too. Simply put, variation in urbanization within sub-Saharan Africa is not well explained by GDP shares in manufacturing, services, and resource extraction. Perhaps future research will find that the lack of definitive patterns reflects measurement error or outliers in data from sub-Saharan Africa. But if urbanization in sub-Saharan Africa is not a consequence of traditional structural transformation and is not well related to natural resource rents, then what are other possibilities?

Perhaps urbanization in Africa is not so much about the benefits of urban density in Africa but more a consequence of especially low rural productivity and offerings of services. Agricultural productivity in Africa is low by global standards, reflecting low irrigation rates, low fertilizer usage, and an attachment to old seed technologies (Ray et al. 2012; Sánchez 2010). Cereal yields in sub-Saharan Africa are half those of South Asia, which in turn are half those of high-income countries, and well below East Asia and Latin America (Henderson and Kriticos 2018). Low rural productivity helps to explain why urban incomes are comparatively so much higher than rural incomes in Africa, conditional on education, age, gender, and the like (Henderson, Kriticos, and Nigmatulina 2019).

A seeming oddity is that sub-Saharan cities house a surprising number of farmers. Table 2 reports results for a set of twelve countries with a total population of 220 million for which there is relevant data in the Integrated Public Use Microdata Series (IPUMS). The first row shows for different spatial entities, the fraction of workers who report the industry in which they primarily work as agriculture, while row 2 does the same for manufacturing. Thus, for example, in row 2 and column 2, less than 2 percent of workers living in rural areas report the main industry in which they work as being manufacturing. In columns 1 and 2, we report these fractions for all workers living in census-defined urban versus rural areas. The remaining columns isolate the primate (largest) cities in each country, and then those in the top 25 percent by size within each country (excluding the primate),

Table 2
Farmers in African Cities by City Size

Spatial scale	*All urban*	*All rural*	*Primate city*	*Secondary cities (top 25%)*	*Tertiary cities (50–75%)*	*All others*
Percentage of workers reporting agriculture as main industry	20.5	88	8.5	23.8	38.6	41.3
Percentage of workers reporting manufacturing as main industry	10.6	<2	12.4	10	8.3	7.3

Source: Henderson and Kriticos (2018) Figure 3 and Supplemental Figure 2.
Note: Data from IPUMS for the most recent census for Ethiopia, Tanzania, Uganda, Mozambique, Ghana, Cameroon, Mali, Malawi, Zambia, Sierra Leone, Liberia, and Botswana. Small cities are in the bottom 50 percent of cities by size and tertiary cities are in the 50–75th percentiles. Cities are defined by night-light boundaries to which population is assigned.

and those in the 25-50 percent, 50-75 percent, and the bottom 50 percent by city size. The share of agriculture in city employment by city size type ranges from 9 to 41 percent and averages 20.5 percent. In fact, in the bottom 75 percent of cities by size, the share of agriculture in urban employment averages 40 percent in this sample. In contrast, in Brazil, India, and Malaysia, shares of urban farmers are all under 7.5 percent. Table 2 also shows that in these sub-Saharan African countries, 88 percent of rural sector employment is in farming. This is far higher than other countries, where rural services, construction, and even manufacturing employment are more important. Finally, note the especially small manufacturing share in smaller cities and towns. Most likely, any manufacturing in these places is traditional food processing, non-metallic minerals, locally made furniture, weaving, and the like for local consumption.

The table tells us African cities are home to a substantial number of farmers. Why do farmers move to cities and by inference commute out to farms? One answer may be better access to amenities and public services as well as consumer services. This in turn may be related to the absence of almost anything but farming in rural areas, which may also reflect a lack of rural infrastructure and institutions. Another answer may be the better employment opportunities in cities for other family members, both in terms of hours worked and the diversity of occupations available (Henderson et al. 2019). Moreover, the large number of urban residents who report their primary occupation as farming may also work in the off-season in other occupations.

Apart from so many farmers living in cities, there is a literature suggesting that Africa may bypass the development of modern manufacturing. The papers in Newfarmer, Page, and Tarp (2018) suggest that African development may rely more on tourism and aspects of information technologies as well as work related to farming such as food processing and horticulture. Henderson and Kriticos (2018) show for a sample of five countries that tradable urban services, like finance, are

growing at extraordinary rates, albeit from a very low base. With that being said, evidence for the current level and trajectories of urban employment by sector is fragmentary. Understanding how sub-Saharan Africa is urbanizing remains a subject of debate, and one that would benefit from more and better data.

Density and Population

Up to this point, our discussion has used definitions of *urban* that are based on host-country specific definitions and implemented using data that may also vary qualitatively from country to country. Unsurprisingly, these definitions are not consistent across countries and may involve subjective assessments like whether an area has certain public facilities, administrative responsibilities, or has a central economic core. Moreover, the extent of metropolitan areas is typically based on the boundaries of country-specific administrative units (like counties in a US context). In some cases, national definitions, especially for capital cities, have tended to severely restrict official urban area size on the basis of historical criteria like a defined national capital zone (as is the case for Jakarta).

One way to avoid such classification problems is to focus instead on population density. Density can be used to define urban areas based on density cut-off points as for example in the Global Human Settlement data (GHS-SMOD L1). However rather than use arbitrary cut-off points and, as noted above, to be consistent with modern modelling, we allow density to vary continuously across space. This is also in line with evidence for developing counties that agglomeration economies arise from density rather than absolute labor market size (Chauvin et al. 2017; Combes et al. 2020; Quintero and Roberts 2018; Henderson, Kriticos, and Nigmatulina 2019).

Which data sets give us finely gridded densities? Perhaps the best-known is the Gridded Population of the World version 4 (GPWv4; CIESIN 2018), which uses population data for country-specific administrative or enumeration units used in their census. GPW sets up the world in grid cells of (approximately) one kilometer. However, GPW then has to map the census unit data into these grid squares, where census enumeration units may be larger or smaller than these grid cells. The census units for which data are released may be quite large administrative units, such as a county or even a province. In these cases, GPW prorates enumeration unit population to grid cells by assuming population is spread uniformly over each reported unit. While high-income countries like the United States often release population data on a fine spatial scale, that is not the case in developing countries. For example, of the 12.9 million polygon-shaped administrative units that form the basis for population estimates in the global GPW, only 2.4 million are from outside the United States.

Rather than use the GPW, we use the European Union's Global Human Settlements population layer (GHS-POP from Schiavina, Freire, and MacManus 2019; Freire et al. 2016). The GHSL again allocates GPWv4 population estimates across one-kilometer grids, but instead of assuming that population is evenly distributed across a polygon-shaped enumeration area, it allocates people according to the

spatial distribution of the footprint of built cover within each area. "Built cover" is based on the EU's specific processing of Landsat data 30-meter resolution satellite data circa 2015 (Corbane et al. 2018, 2019).[2] In the rare cases where there is no built cover in an enumeration polygon, it reverts to the GPWv4 estimates. More information about the GHS data can be found in Florczyk et al. (2019).[3]

Based on these data, we present information about population density per square kilometer for grid squares whose size is one square kilometer at the equator. We compare Europe, North America, sub-Saharan Africa, Latin America, and South Asia. We pool East and Southeast Asia together to improve the legibility of our figures.

In Figure 1a, we graph the cumulated share of population by density. Clearly, North America and then Europe have the highest accumulated shares of population at low densities. In America and Europe, less than 10 percent of the population lives at densities above 10,000 people/square kilometer. Sub-Saharan Africa, along with East and South East Asia, have the lowest accumulated shares at low density, or equivalently, the highest degree of density inequality. In sub-Saharan Africa and in East and Southeast Asia, 30-40 percent of the population lives at densities above the 10,000 threshold, while for Latin America and South Asia, it is about 20 percent.

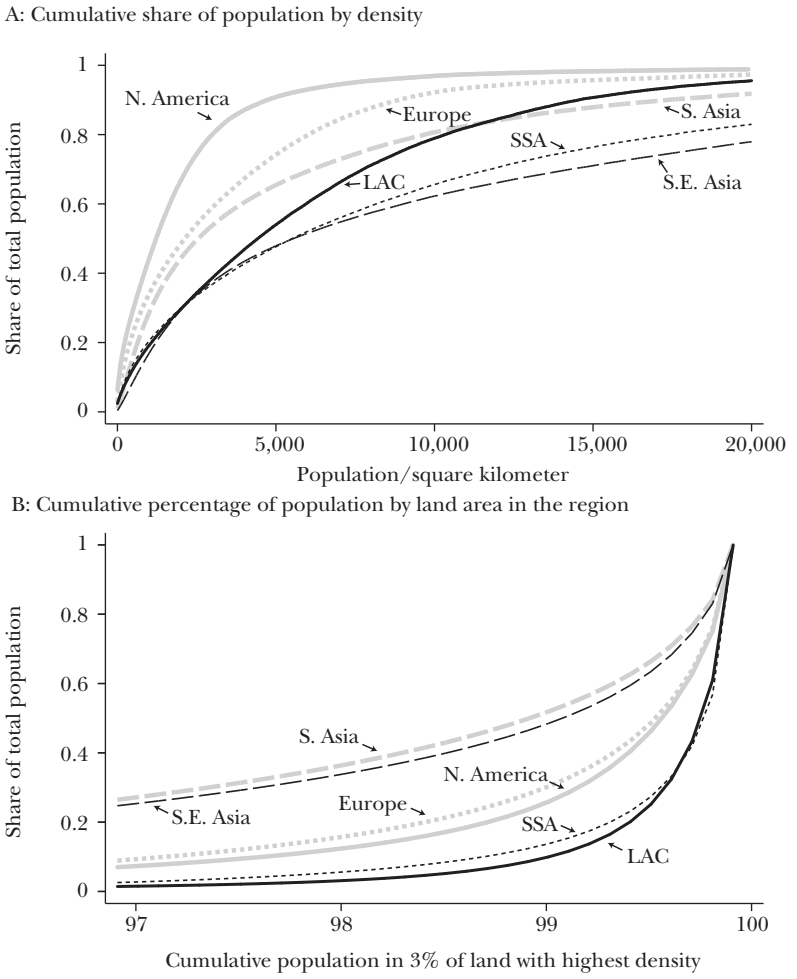
To improve legibility, our graphs stop at 20,000 people/square kilometer. In Southeast and East Asia, 18-20 percent of the population lives at densities above 20,000/square kilometer. In the developed world, the proportion of people living at such densities is tiny. For the purpose of understanding density and its implications, the developed world probably cannot teach us much about the very high densities experienced by a significant portion of the developing world's population.

Figure 1b graphs the corresponding cumulated share of population by area. It shows only the 3 percent of regional area that is most densely populated. Looking

[2]Landscan data is an alternative source of gridded population data. These data rely on a proprietary algorithm to construct population estimates based on higher resolution satellite imagery than Landsat and information on airports and rails (Rose and Bright 2014). The algorithm is not publicly documented and changes from year to year. Moreover, the estimates are for ambient population averaged throughout the day, whereas GHS-POP is for the nighttime (residential) population. We choose the GHS data because it is consistently defined over time and the algorithm is public. One issue concerns how all these data sets deal with the vast number of grid squares with very low or zero population worldwide. The GHSL, Landscan, and GPW handle the problem very differently. However, the densities we look at for our purposes (say, above 8 people per square kilometer), based on other work in progress, the distributions are quite close.

[3]As noted in the text, the GHS Settlement Model (GHS-SMOD L1) also attempts to define city status based on density and population cut-offs. Starting from gridded population data, "cities" are defined as sets of contiguous one-kilometer grid cells having density over 1,500 people and summing to over 50,000 people. The settlement model also constructs "towns" and suburbs," defined as sets of contiguous pixels with density and size thresholds of 300 people per square kilometer and 5,000 in total. This approach has the advantage of avoiding administrative boundaries for classifying urban areas, but it also seems arbitrary. For example, it would be hard to argue that an agglomeration of people satisfying such a definition would accurately describe the actual labor market or commuting zone of a city. Indeed, appropriate density cut-offs would probably vary by country and region of the world. Our discussion in the text focuses only on density and ignores the definitions that would emerge from using these population cut-offs.

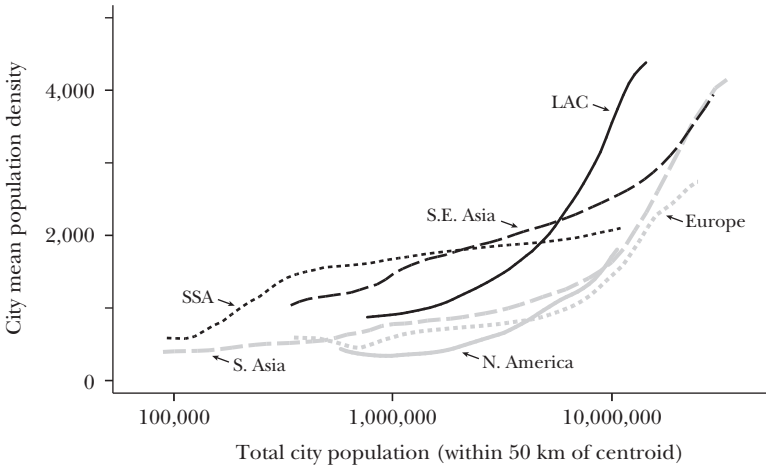
Figure 1
Population Density Gradients by Region



Source: Based on population data from GHS.
Note: SSA—sub-Saharan Africa, LAC—Latin America and the Caribbean.

at the y-axis in this figure, we see that about 25 percent of the population of South and South East Asia occupies the 97 percent of regional area that is the least densely populated, while in other regions that population share is small—especially in North America and Europe. As a result, almost everyone lives on less than 3 percent of the land area. The more widespread occupation of land in South and South-East Asia reflects two factors: 1) a larger fraction of Asian land employed in labor-intensive agriculture; and 2) in much of Asia there is a relatively high ratio of national population to land area, forcing use of a greater proportion of land. But even in South and South East Asia, there is still a lot of room for people to live at lower density: just 25 percent of the population occupies 97 percent of the land area.

Figure 2
City Population Density by Total City Population



Source: Based on population data from GHS.
Note: Vertical axis is mean population density from GHS in a 50km radius disk centered on the centroid of each of the 657 UN world cities. Horizontal axis is total population in the same disk.

Combining the results of the two parts of the figure, we note that North Americans live at relatively low densities but are endowed with a large land area, so most land is very sparsely populated. In contrast, many Africans live at high densities as we saw in Figure 1a, while in Figure 1b, most land is also sparsely populated. There are many other factors apart from regional land availability per person determining the patterns of population density and land use that we see in Figures 1a and b.

Figure 2 describes the relationship between city size and urban population density but with a common geographic measure of a "city." To construct Figure 2, we use the 657 cities described by the UN World Cities data. These are cities that housed more than 300,000 people at any time between 1950-2010. For each city, the UN World Cities data reports the latitude and longitude of the center of the city. We draw a 50-kilometer radius around each such centroid and sum the population in the gridded population squares within this disc. To measure population density for these cities, we calculate the person-weighted density of grid cells within each city's disk. Figure 2 presents local polynomial regressions of the relationship between city population and density by region.

In most regions of the world, higher densities are associated with larger city populations. This pattern is clear in North America. However, in Africa the relationship is weak. For most regions of the world, mean population density in the 50-kilometer disk rises quickly with total population above about 450,000, although in North America, the take-off point is near 750,000. Africa differs. Below 450,000 and above about 2 million, African cities have higher densities than in other regions. As city sizes increase in Africa, density rises relatively slowly.

Evidence on How Living Conditions in Developing Countries Vary across Density

When studying agglomeration economies and diseconomies, with a typical focus on cities, researchers have employed different scale measures such as total city employment or population or a measure of density. When looking at total city population, the researcher is largely constrained to accept an administrative or other boundary of an urban area, which then makes an implicit assumption that any resulting agglomeration economies operate uniformly within that boundary. Here, we will instead consider how a variety of outcomes vary continuously with population density. This not only allows for within urban area variation but treats space as a continuum (as in Michaels, Rauch, and Redding 2012; Desmet and Rappaport 2017).

Our empirical methodology is straightforward. To learn more about how peoples' lives change with population density, we measure density using the Global Human Settlements data described in the previous section. For measures of outcomes, we turn to three sources noted in the introduction: the Living Standards and Measurement Survey for data on income and wages in six sub-Saharan African countries; the Demographic and Health Surveys for data on female outcomes, child outcomes, infant mortality, household utilities, schooling, and adult lifestyle outcomes for 40 countries in Latin America, South East Asia, sub-Saharan Africa and South Asia (with a focus on the last two regions), generally conducted from 2010–2016; and Afrobarometer for data on crime in 24 sub-Saharan African countries. Details on the DHS and Afrobarometer are in Henderson et al. (2020). As is common, these surveys use a "cluster" approach of questioning randomly selected people near a smaller number of randomly selected "cluster" points and assign all such respondents the location of the cluster.[4] With our gridded data, we can draw a five-kilometer disc surrounding each cluster, and in this way, we can match geocoded individual-level surveys to population density. Thus, we are able to examine how survey responses describing income, health, education, public health, and public goods vary with nearby density in a large sample of developing world countries.[5]

[4] For the Afrobarometer and Living Standards and Measurement Survey, clusters are generally located at the centroid of a small administrative unit, such as the finest census enumeration unit. To protect respondent privacy, clusters in the Demographic and Health Surveys are displaced by up to two kilometers for urban respondents and up to five kilometers for rural respondents. This introduces some error into the respondent relevant measure of population density. We truncate respondents in five-kilometer radiuses with population densities less than 7.4 people per square kilometer because we are suspicious of the accuracy of GHSL estimates at low population densities. We also observe dramatically wider confidence bands around non-linear regressions of outcomes on log density below this threshold.

[5] Our approach is conceptually similar to that of Gollin, Kirchberger, and Lagakos (2017), who look at the relationship between various outcomes reported by the Demographic and Health Surveys and population density in an area around clustered survey respondents. They find that survey respondents living at the 80th or 90th percentile of the set of DHS cluster densities typically have better amenities than those for people living at the 20th or 10th percentile. These results are interesting and important, but somewhat difficult to interpret. As we saw earlier in Figures 1a and 1b, population is highly concentrated into very small, very dense regions in sub-Saharan Africa and South and South East Asia. The 80th or

To illustrate our results, we focus on figures constructed using the "binscatter" methodology described in Cattaneo et al. (2019). In our figures, we show the outcome for an (endogenous) number of equal size bins. The confidence bands describe the region around local polynomial regressions in which we expect a local polynomial regression line to lie with 95 percent probability. In each figure for the left-panel non-parametric estimates, we do not include control variables. In the right-panel semi-parametric regressions, we include country fixed effects and a range of control variables, which differ somewhat by outcome according to various factors like whether the outcome in question reflects a household, person, or child-level outcome, or what was included in the survey instrument. Broadly, the control variables reflect the education, gender, and age of the household, person, or child whom the survey response describes.

The figures also report a line of best fit and its slope coefficient. If this best-fit line falls outside the confidence band from the binscatter, then a linear relationship can be rejected—at least locally. While generally the regression lines lie within the confidence bands, the illustrative graphs on which we report show very different widths for confidence bands. In an online Data Appendix available with this paper at the journal's website, we offer a detailed presentation of ordinary least squares regressions, with and without control variables, as well as the list of specific controls, the countries, and the like.[6] We note that the density gradients we report can be based on quite different samples of countries, and so some caution is required in comparing regression results across outcomes.

This methodology has well-known weaknesses, and the evidence presented here should be viewed as a suggestive first pass in analyzing multiple data sets, which deserve more in-depth work. Our results are associations and do not give magnitudes of true causal effects. For example, while we will find a rapid rise in income with density even with control variables, omitted variables such as ability and ambition are surely also important, and may influence how people sort across rural and urban locations. Part of the association of higher incomes with density could be that, conditional on education, higher ability people may live at higher densities. Of course, higher ability people may benefit more from higher densities, so density effects are heterogenous. As we will note below, controlling for education in income regressions may lead us to understate some benefits of density for those with lower incomes, in terms of facilitating better schooling.

Resolving these inference problems is difficult and beyond the scope of this project. A few experiments have sought to induce random variation in subject

even 90th percentile of DHS clusters by density is not very dense, especially given the enormous rural over-sampling in the DHS. Implicitly, the Gollin, Kirchberger, and Lagakos (2017) methodology is telling us about the distribution of amenities across places rather than across people according to how they live. Given the small proportion of the landscape occupied by cities, it is hard to interpret these findings in terms of a difference between the rural and urban experience.

[6]We note the similarity between results presented here and those in Henderson et al. (2020), which examines how outcomes differ across the discrete urban-rural classifications given in the GHS Settlement Model (GHS-SMOD L1) noted earlier.

locations, but most are within city or involve refugee and other programs applied to very special populations (reviewed in Bryan, Glaeser, and Tsivanidis 2019). Extending these experimental and quasi-experimental methods to the larger set of outcomes that we consider is an obvious area for further research.

Incomes

The LSMS provides survey data on income and wage (for hourly workers) for six African countries—Ethiopia, Ghana, Malawi, Nigeria, Tanzania, and Uganda—with a combined population of over 400 million people. Household income is constructed from several LSMS questions. It includes all wage income and business receipts (including farm), less business expenses per month (for details of variable definitions, see Henderson and Kriticos 2018).

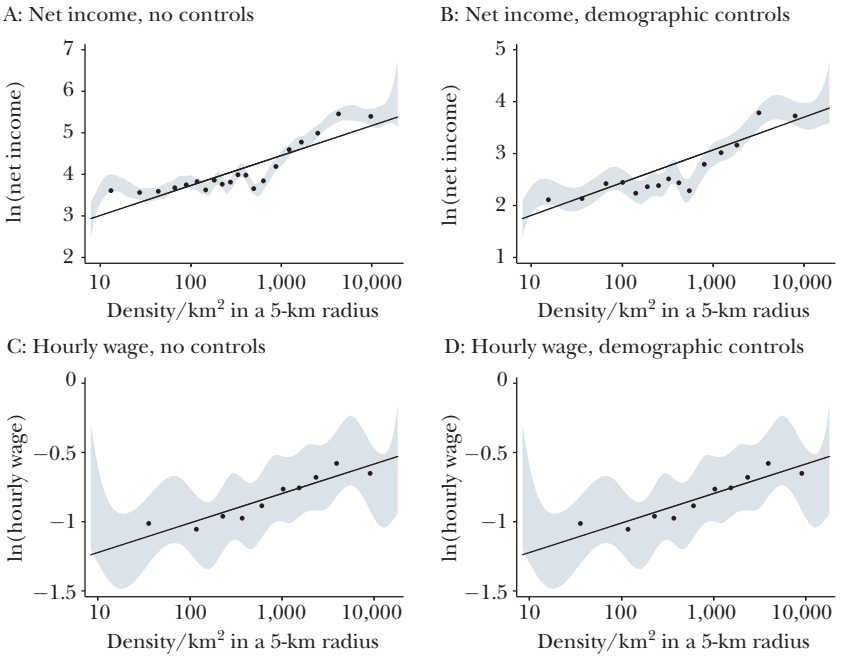
The top two graphs of Figure 3 show the binscatter plot of (the log of) net household income against (the log) of density, with and without control variables. The bottom two graphs show a similar plot for wage data. The estimated elasticities from the best-fit line reported in the figure are high. Doubling density increases household net income by about 32 percent and hourly wages by about 5 percent, with controls. At 5 percent, the density elasticity of hourly wages exceeds those typically found in developed countries but is in the range of estimates in recent work on other developing regions and countries (for example, Quintero and Roberts 2018; Duranton 2016; Combes et al. 2020; Chauvin et al. 2017). Yet the income elasticity is probably more important. After all, it is families that migrate permanently to cities. The fact that the density elasticity of net income is a multiple of that for wages likely reflects both an increase in hours worked and varieties of job opportunities for family members, as analyzed in Henderson et al. (2019). We know of no comparison for the density elasticity of net income in the literature.

While the figures suggest that a linear fit is reasonable, the graph also suggests a potential non-linearity. The density gradient is flatter from about 8 to 550 people/square kilometer. This is well below the average density of African cities (shown earlier in Figure 2). We do not think this flat portion has to do with measurement of income. Incomes appear to be measured as well in this low-density part of the graph as other parts, given the detail and high standards of the LSMS. After this point, the gradient increases sharply such that a household moving from a density of 550 to 8,100 people/square kilometer shows about a four-fold increase in income. While the LSMS reports respondents at densities near 20,000 people/square kilometer, such respondents are rare and so our estimates of income at these high densities are imprecise. The corresponding plots for hourly wages are similar but with a less steep slope and modestly wider confidence bands. In all, these estimations indicate that African wages and income are sensitive to density and suggest that moving to denser locations may have a high return for African families.

Utilities and Schooling: Public Goods Strongly Influenced by Policy

Household access to utilities and schooling depend in greater part on public sector provision of, for example, water mains, reservoirs, schools and teachers. The

Figure 3
Log of Household Net Income and Hourly Wage versus log Population Density



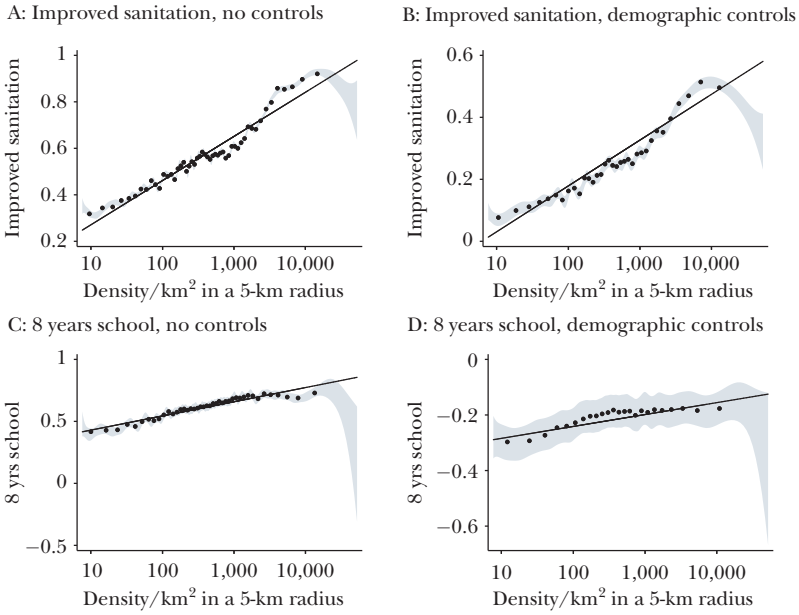
Note: Binscatter plots of LSMS net income of respondent household and of hourly wage, against the log of GHS population density in a 5km disk around the survey respondent. Log population density is censored below at about 8/km^2. Left panels have no controls. Right panels includes demographic controls and country fixed effects. Shading indicates 95 percent confidence band. Income includes wage income, net farm income, and net business income. For a small number of observations expenses exceed (monthly) incomes. We drop these observations to permit logarithmic scaling. LSMS survey countries are listed in table A2. Linear regression based on results in table A1a, which provides more details about the sample. Slope coefficients and standard errors of best linear fits are; (a) 0.313 (0.016) (b) 0.317 (0.014) (c) 0.118 (0.015) (d) 0.049 (0.009). Details in online Appendix.

Demographic and Health Surveys ask questions about electricity, safe drinking water, improved sanitation, and educational attainment. The questions about water and sanitation are nuanced and tailored to allow an evaluation of whether the UN's sustainable development goals are attained. For example, "safe" water can be quite different than piped water. In cities in sub-Saharan Africa and in South Asia (as defined in the GHS settlement model) about 40 and 80 percent of people have access to "safe" water, but only about 8 and 25 percent respectively have water piped into their dwelling unit (Henderson et al. 2020). Toilets flushing into a central sewer system are rare in these cities.

The top two graphs of Figure 4 present a scatterplot plot where the outcome variable is the indicator for improved sanitation. Even after controlling for household demographic characteristics, we see a rapid and precisely estimated increase in

Figure 4

Access to Improved Sanitation and Probability of Children Receiving Eight Years of School versus log Population



Note: Binscatter plots of a DHS indicator variable that is one if a respondent household has access to improved sanitation and of an indicator that is one if a household child 16 years old completed eight years of school, against the log of GHS population density in a 5km disk around the survey respondent. Log population density is censored below at about 8/km^2. Left panel is unconditional. Right panel includes demographic controls and country fixed effects. Shading indicates 95 percent confidence band. DHS survey countries are listed in table A2. Linear regression based on results in table A1a, which provides more details about the sample. Slope coefficients and standard errors of best linear fits are; (a) 0.083 (0.001) (b)0.063 (0.001) (c) 0.050 (0.001) (d) 0.016 (0.001). Details in online Appendix.

access to improved sanitation with density. As in the earlier case of net income, we also see a slow increase in access to improved sanitation at lower densities and more rapid increase at higher densities. Going from 550 people to 8,100 people/square kilometer raises the likelihood of improved sanitation from under 25 percent to over 50 percent. There is also evidence of a downturn at very high densities. This may reflect a decline in services to high-density slums, but our limited sample of very high-density respondents does not allow precision at this tail.

A figure for access to safe water looks similar, including the non-linearity at high densities. For electricity, the fit with control variables included is very tight, although the rise is more linear. With mean outcomes of 0.5 to 0.7 for these three utilities, a one-standard deviation increase in the log of density increases outcomes from 0.075 to 0.11. We believe these differences are supply-driven and reflect lower

costs of service provision in dense areas, perhaps along with political considerations. Denser areas may be more favored in the political arena as in the classic urban bias literature (for example, Ades and Glaeser 1995; Davis and Henderson 2003).

The bottom two graphs of Figure 4 show a relationship between density and schooling. In our estimates, the schooling outcome is for 16-year-olds and is an indicator variable that takes value one in the event that a household 16-year-old has completed at least 8 years of schooling, and zero otherwise. For the best-fit line in the left-hand figure, increasing density by one log point increases the share of 16-year-olds completing eight years of schooling by about 0.050. Controls reduce this effect by two-thirds to 0.016 in the right-hand figure, so that a one standard deviation increase in ln(density) raises the probability by 0.027 for a sample mean of 0.61.

Density effects for schooling are smaller than utilities. However, why should schooling attainment, after controlling for family characteristics, be affected at all by density at all? One possibility worth exploring is a more reliable supply of schooling and teachers in denser areas.

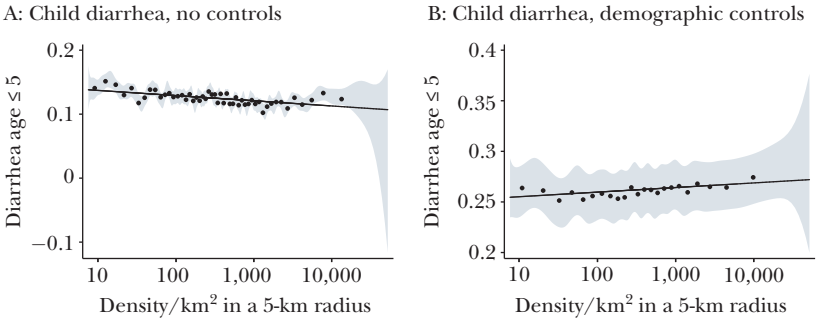
This raises an important issue. In examining the income and wage returns to density, we tried to control for sorting by controlling for education. However if higher density in developing world cities has a causal effect on human capital accumulation for families who move to density, education ought to be seen as part of the benefit of density, not as a sorting effect that should be held constant when looking at higher urban incomes. How to separate out these components is a subject for future research.

Female, Child, and Birth Outcomes

The DHS also reports on a variety of indicators related to the status and well-being of women and children. For example, indicator variables include: the use of modern contraception for sexually active women ages 20-40 who are not pregnant and do not want to have a child in the next two years; reporting an affirmative response by females to the question "is wife beating justified for any reason?"; and if a woman reports having ever experienced spousal household violence. The data also include the total number of births in the last three years to each woman age 15–49, and whether each child born from three months to three years ago survived at least three months. For each household child, there are indicator variables reporting whether the child has had the third and final DPT3 immunization shot by two years of age, whether the child has had diarrhea in the last two weeks, and whether each child age five and under has had a cough in the last two weeks.

As one illustration of the general findings, Figure 5 presents a binscatter plot of the relationship between the incidence of childhood diarrhea and density. First, as in this figure, best-fit lines for these outcomes indicate small marginal effects of density, but generally, incidence rates are also low. For example, here the (significant) slope in the left-hand graph is −0.0035 for an average incidence of 0.125. Second, all unconditional outcomes improve as density rises, except for being a victim of spousal abuse and whether a child five-or-under has had a cough recently.

Figure 5
Diarrhea Last Two Weeks for Children Five and under versus log Population Density



Note: Binscatter plots of a DHS indicator that is one if a child five or under had diarrhea in the past two weeks, against the log of GHS population density in a 5km disk around the survey respondent. Log population density is censored below at about 8/km^2. Left panel is unconditional. Right panel includes demographic controls and country fixed effects. Shading indicates 95 percent confidence band. DHS survey countries are listed in table A2. Linear regression based on results in table A1b, which provides more details about the sample. Slope coecients and standard errors of best linear fits are; (a) –0.004 (0.0005) (b) 0.003 (0.0004). Details in online Appendix.

Third and most critically, using control variables changes the picture considerably. In a number of cases, demographic controls reduce density coefficients by well over 50 percent. However, most critically, in the case of diarrhea as illustrated, along with cough and infant mortality, effects are actually reversed, and being a victim of spousal abuse, having diarrhea, having cough and infant mortality increase significantly with density once controls are added. After controlling for demographic characteristics, one standard deviation increase in density is associated with an increase in domestic violence, diarrhea, cough, and infant mortality of 3.5 to 5 percent relative to their means.

Finally, with the addition of controls, the confidence bands expand dramatically, as illustrated in the right-hand panel in Figure 5. This huge widening of confidence bands once controls are added applies to most of the outcomes in this sub-section (with the exception of fertility and spousal abuse). This means that we can have less confidence in the local precision of marginal density effects for most outcomes discussed in this section despite significant slopes to best-fit lines. To put it another way, the relationships among density, demographic controls, and these outcomes need much more investigation and may be more subject to unobserved features of the local environment.

The finding that diarrhea may rise with density may seem at odds with the finding in the previous subsection that that safe water and improved sanitation both improve with density. One possible interpretation is that, as density rises, the increased access to safe water and improved sanitation is not enough to offset the effect of increased crowding on contamination of food and water. Another

possibility is that the UN sustainable development goals are setting too low a bar: what is being counted as safely managed water and improved sanitation is not clean enough.

Lifestyle Diseases and Crime

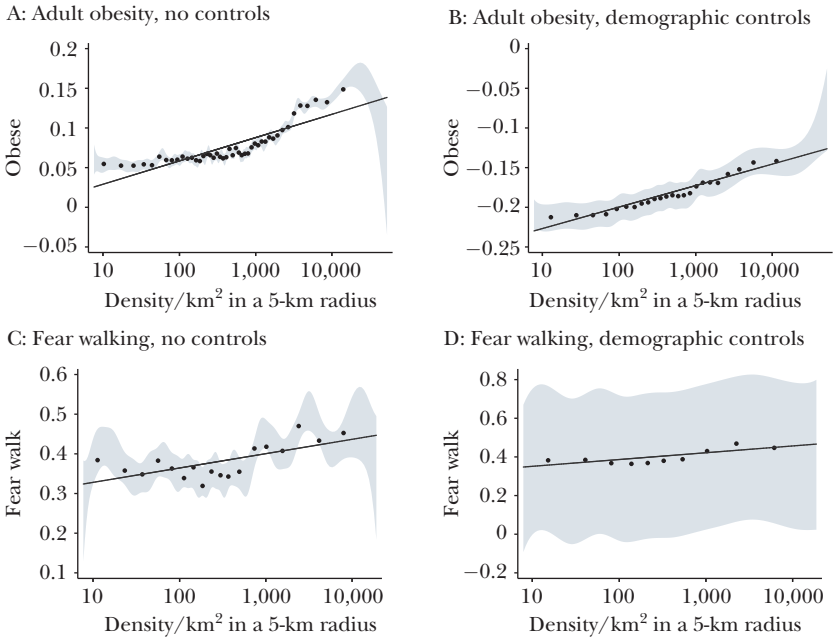
The DHS data allows us to report on the relationship between population density and four lifestyle diseases for adults age 20–49. Obesity data exist for all our countries. For India and Nepal, there are data on the incidence of measured high blood pressure, and for India alone, there are data on self-reported asthma and diabetes. These are to some extent lifestyle diseases. They reflect at least in part patterns of diet, exercise, work intensity, and stress, which may be affected by higher density. We can imagine that stress might come from long commutes or hours worked, smaller social networks, changed family circumstances, and crowding.

The top panel of Figure 6 shows the scatterplots for obesity, which is defined here as having a Body Mass Index (BMI) above 30. (The formula for BMI is weight in kilograms divided by height in meters squared.) The incidence of obesity, high blood pressure and diabetes all increase with density. Controls have a small impact on marginal outcomes, either raising or lowering them by less than by 25 percent. In the right-hand graph, we see that the slope of the best-fit line is 0.010 for a mean of 0.077, so that an increase of one standard deviation in density (1.7) is associated with a 22 percent increase in obesity from the mean.

We note that asthma does not respond to density. This perhaps surprising result would seem to be consistent with results in Aldeco, Barrage, and Turner (2019). Using global data, that paper finds the relationship between population density and the concentration of airborne particulates to be unambiguously positive but quite small: that is, air pollution is worse in cities than in rural areas, but not much worse.

The Afrobarometer survey collects data from 26 African countries about four sets of feelings or outcomes about crime: whether the survey respondent reported being fearful while walking outside in their neighborhood; whether the respondent reported being fearful of crime at home; whether the survey respondent's home has been robbed in the past year; and whether anyone in the household has been attacked by an outsider in the past year. Results are fairly similar for all these outcomes, in terms of marginal density effects relative to average incidence with effects rising with density. For illustration, the bottom panel of Figure 6 shows a binscatterplot for being fearful while walking outside, where a one standard deviation (1.8) increase in density is associated with a 0.029 increase in fear, for an average incidence of 0.38. While the left-hand graph suggests a sharp rise in fear in the higher density ranges, in the right-hand graph, the confidence intervals on local marginal effects with controls are very large. Finally, we note that, for actually being attacked in the past year, the slope of the best-fit line is insignificant and the incidence at 0.10 is much smaller than those for other outcomes including, from above the fear of being attacked. Perhaps those with a greater fear take more precautions to avoid actual incidence.

Figure 6
Adult Obesity and Self-Reported Fear of Walking Outside versus log Population Density



Note: Binscatter plots of a DHS indicator that is one if the survey respondent is obese or reported being afraid for their safety while walking outside, against the log of GHS population density in a 5km disk around the survey respondent. Log population density is censored below at about $8/km^2$. Left panel is unconditional. Right panel includes demographic controls and country fixed effects. Shading indicates 95 percent confidence band. DHS survey countries are listed in table A2. Linear regression based on results in table A1b, which provides more details about the sample. Slope coefficients and standard errors of best linear fits are; (a) 0.013 (0.0005) (b) 0.010 (0.0003) (c) 0.016 (0.004) (d) 0.016 (0.003). Details in online Appendix.

Summary

While incomes, wages, access to utilities, schooling, number of births, and use of contraception and vaccinations all improve with density, we see declines with density in child and adult health outcomes, including infant mortality, obesity, domestic violence, and fear of crime. Of course, some people will worry more about lifestyle diseases or crime than others, although child health may be harder to put aside. Regardless, those who place a heavier weight on these factors may be less likely to migrate to urban areas.

The Roback Model Meets Current Patterns of Urbanization

The Roback (1982) model is the workhorse model for thinking about spatial equilibrium. In the original model, people are identical and move across space to

equalize utility levels. An important innovation in the recent literature has been to introduce moving costs and different forms of individual heterogeneity. With these additions, agents no longer move across locations to equalize utility levels. Rather, all agents choose their favorite location, taking into account the cost of moving there from their starting point. To illustrate, suppose a continuum of people choose between an urban and rural location. In the most general set-up, people receive a location and individual specific income, an "amenity" that is valued similarly by all agents (for example, up to income effects), and an "affinity" draw that is person- and location-specific. Amenities represent location specific attributes like the availability of safe water, the prevalence of crime, or the difficulty of commuting. Affinities reflect things like a personal taste for local weather or landscape or the presence of family members or roots in a home location.[7] Finally, to move between locations, agents must pay a migration cost.

In a static spatial equilibrium, no one can gain by moving, at least not after accounting for the costs of moving. The notion of spatial equilibrium provides us with a powerful framework for organizing our ideas about what causes different people to arrange themselves across the landscape in the ways that we observe. At its heart, the model assumes that people act to arbitrage spatial differences in productivity and amenities by changing locations. Their ability to conduct such moves is hampered by frictions, moving costs, and idiosyncratic attachment to locations.

Consider the simple case in Moretti (2010), where there are no moving costs. All people *within* a region receive identical real incomes and amenities, and at the margin, real incomes are declining in population in each of the two regions. Then, there is a marginal person whose affinity draws make that person exactly indifferent between living in the two regions. For example, in the urban region, everyone has a weaker relative affinity for *the rural region* than the marginal agent who chooses the urban region. Note that in the resulting equilibrium, utility levels are not equalized across agents nor are real incomes equalized across regions, except in special cases.[8]

To illustrate these ideas, we have stated the model in a very simple form. We suspect that people are "more biased" towards the place they are born. To accommodate this, some formulations shift the distribution of affinity draws for the "birth location" to the right of the other locations. While this is intuitively appealing, it is practically similar to a change in moving costs in our formulation. While we assumed that moving costs are the same across people and independent of the direction of the move, this assumption is clearly incorrect in many contexts. In China, for example,

[7] In practice, these different draws are typically imagined to arise from an econometrically convenient distribution, most often an extreme value distribution.

[8] If the two regions offer identical amenities, endowments, and technologies, agents are identical absent their affinity draws, and the distribution of the differences in affinity draws is symmetrical about zero, then real incomes will be equalized and the marginal person will have equalized affinity draws. However, if the urban region has superior endowments or technologies, then we generally expect an equilibrium where real incomes are higher in the urban region and the marginal person has a greater affinity draw in the rural than urban location.

one would expect moving costs to vary on the basis of *hukou* status (whether a person is registered to live as a citizen in an urban or a rural place) and the direction of the move.

Note that without restrictions on moving costs or idiosyncratic affinities, the model has no content. If moving costs are sufficiently high, we can rationalize any observed outcome. People could be like trees: they stay where they are planted no matter how much their wages might increase if they moved over the hill. Similarly, we can always choose affinity draws such that everyone will want to stay where they are born.

We are just beginning to learn about the importance for mobility of frictions like moving costs and affinities. In a static model, Tombe and Zhu (2019) find enormous moving costs for Chinese migrants from rural farms to urban factories. Moving within a province costs migrants over half of their real income (at destination) and moving across provinces increases this to more than 90 percent. Similarly, Bryan and Morten (2019) on Indonesia argue that moving 1,000 kilometers from the place of birth costs 40 percent of real income and moving 200 kilometers costs 20 percent. Note that these two papers, like ours, rely on a static model, while "migration" is explicitly a dynamic concept. Working with models of spatial equilibrium with dynamics is difficult but is an active area of research (for example, Balboni 2019; Caliendo, Dvorkin, and Parro 2019; Ahlfeldt et al. 2020). Finally, there are other considerations of incomplete markets and risk raised by the experiments conducted in Bangladesh in the context of round trip or seasonal migration, which also find high disutility or lack of affinity from migration, as reviewed in Lagakos, Mobarak, and Waugh (2018).

Putting aside the relative lack of empirical evidence on attachment and moving costs, if we are willing to assume that moving costs are not too large and differences in affinities are limited, then we should not observe the case where both amenities and incomes are much higher in one populated place than another.

The theoretical model thus suggests that the "puzzle of early urbanization" might be rephrased as the "puzzle of too-slow urbanization." Recent empirical work and our own results indicate wages and household incomes in developing world cities are dramatically higher than in the countryside, even after we condition on individual age, gender, and education. Moreover, the data show clearly that access to safe water, electricity, and modern sanitation improve rapidly with urbanization.

These observed patterns suggest that something is slowing down a faster pattern of urbanization that would otherwise be happening: in other words, it suggests that mobility costs and spatial attachment matter. That said, our more exhaustive accounting suggests that, while much about urban life is better than rural life, at least some things are worse, such as adult and child health outcomes and crime. Therefore, together with mobility costs and spatial attachment, if people also trade off the costs and benefits of urban life at plausible rates, current rates of urbanization in developing countries can be consistent with the spatial arbitrage that is the foundation of models of spatial equilibrium.

Conclusion

The new metropolises of the world are being built in sub-Saharan Africa, South Asia, and South East Asia. However, the mechanism that seems to have driven urbanization in much of the rest of the world—the decline of labor productivity in agriculture relative to manufacturing—may not always be at work. In some regions of the developing world, and in sub-Saharan Africa in particular, people are moving to cities when they are poorer and less productive than were their nineteenth and twentieth century counterparts in developed countries. In addition, population densities in many urban areas of South and South East Asia and sub-Saharan Africa are also much higher than what we observe in developed countries.

We also presented evidence, confirming findings of earlier research, that incomes and wages increase rapidly with density. Moreover, in spite of the "earliness" of developing world urbanization, many important aspects of life improve rapidly with density: access to electricity, safe water, modern sanitation, schooling, and inoculations for children. In seeking to understand how these factors and patterns interact, we turn to a variant of the classic Roback (1982) model of spatial equilibrium. Its basic intuition is that people will move to exploit utility differences across space. However, the benefits of urbanization seem large both economically and econometrically. Against these benefits, the costs of density seem more modest. We consider possible additions to the basic model—like costs of moving or affinities for certain traits of urban or rural areas—that might help to explain why rural-to-urban migration in developing countries is not even higher than we currently observe. Finally, our results suggest that reductions in urban crime and public health interventions that target lifestyle diseases, child health outcomes, and crime may be important tools for policymakers who would like to facilitate rural to urban migration.

■ *We are grateful to Geetika Nagpal, Vivian Liu, and Julia Lynn for excellent research assistance on this project. We thank Sebastian Kriticos and Jamila Nigmatulina for their preparation of the wage and income data which was used in the Kriticos and Henderson (2018) and Henderson, Kriticos, and Nigmatulina (2019) published papers and Cong Peng and Vivian Liu for their preparation for the DHS and Afrobarometer data used in the Henderson et al. (2020) European Union report. We thank the editors for all their very useful comments and suggestions, and additionally, Timothy Taylor for all his editorial work.*

References

Ades, Alberto F., and Edward. L. Glaeser. 1995. "Trade and Circuses: Explaining Urban Giants." *Quarterly Journal of Economics,* 110 (1): 195–227

Ahlfeldt, Gabriel M., Fabian Bald, Duncan Roth, and Tobias Seidel. 2020. "The Stationary Spatial Equilibrium with migration Costs." Unpublished.

Akbar, Prottoy A., Victor Couture, Gilles Duranton, Ejaz Ghani, and Adam Storeygard. 2018. "Mobility and Congestion in Urban India." The World Bank.

Aldeco, Lorenzo, Lint Barrage, and Matthew A. Turner. 2019. "Equilibrium Particulate Exposure." Working Paper. Brown University.

Allcott, Hunt, and Daniel Keniston. 2017. "Dutch Disease or Agglomeration? The Local Economic Effects of Natural Resource Booms in Modern America." *The Review of Economic Studies* 85 (2): 695–731.

Balboni, Clare Alexandra. 2019. "In Harm's Way? Infrastructure Investments and the Persistence of Coastal Cities." PhD dissertation, The London School of Economics and Political Science.

Bolt, Jutta, Marcel Timmer, and Jan Luiten van Zanden. 2014. "GDP per Capita since 1820." In *How Was Life?: Global Well-being since 1820*, edited by Jan Luiten van Zanden, Joerg Baten, Marco Mira d'Ercole, Auke Rijpma, Conal Smith, and Marcel Timmer, 57–72. Paris: OECD Publishing.

Bryan, Gharad, and Melanie Morten. 2019. "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia." *Journal of Political Economy* 127 (5): 2229–68.

Bryan, Gharad, Edward Glaeser, and Nick Tsivanidis. 2019. "Cities in the Developing World." NBER Paper 26390.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. 2019. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock." *Econometrica* 87 (3): 741–835.

Caselli, Francesco, and Wilbur John Coleman II. 2001. "The US Structural Transformation and Regional Convergence: A Reinterpretation." *Journal of Political Economy* 109 (3): 584–616.

Cattaneo, Matias D., Richard K. Crump, Max H. Farrell, and Yingjie Feng. 2019. "On Binscatter." arXiv preprint arXiv:1902.09608. https://arxiv.org/abs/1902.09608. (accessed May 31, 2020).

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. "Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets." NASA Socioeconomic Data and Applications Center (SEDAC). https://sedac.ciesin.columbia.edu/data/collection/gpw-v4/documentation. (accessed May 31, 2020).

Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio. 2017. "What Is Different about Urbanization in Rich and Poor Countries? Cities in Brazil, China, India and the United States." *Journal of Urban Economics* 98: 17–49.

Combes, Pierre-Philippe, Sylvie Démurger, Shi Li, and Jianguo Wang. 2020. "Unequal Migration and Urbanisation Gains in China." *Journal of Development Economics* 142 (January 2020).

Corbane, Christina, Aneta Florczyk, Martino Pesaresi, Panagiotis Politis, Vasileios Syrris. 2018. "GHS-BUILT R2018A - GHS built-up grid, derived from Landsat, multitemporal (1975–1990–2000–2014)." European Commission, Joint Research Centre, JRC Data Catalogue.

Corbane, Christina, Martino Pesaresi, Thomas Kemper, Panagiotis Politis, Aneta J. Florczyk, Vasileios Syrris, Michele Melchiorri, Filip Sabo, and Pierre Soille. 2019. "Automated Global Delineation of Human Settlements from 40 Years of Landsat Satellite Data Archives." *Big Earth Data* 3 (2): 140–69.

Davis, James C., and J. Vernon Henderson. 2003. " Evidence on the Political Economy of the Urbanization Process." *Journal of Urban Economics.* 53 (1): 98–125.

Desmet, Klaus, and J. Vernon Henderson. 2015. "The Geography of Development within Countries." In *Handbook of Regional and Urban Economics*, vol. 5, edited by Duranton, Henderson, and Strange, 1457–1517. Amsterdam: Elsevier.

Desmet, Klaus, and Jordan Rappaport. 2017. "The Settlement of the United States, 1800–2000: The Long Transition towards Gibrat's Law." *Journal of Urban Economics* 98: 50–68.

Duranton, Gilles. 2016. "Determinants of City Growth in Colombia." *Papers in Regional Science* 95 (1): 101–31.

Florczyk, Aneta J., Christina Corban, Daniele Ehrlich, Sergio Manuel Carneiro Freire, Thomas Kemper, Luca Maffenini, Michele Melchiorri et al. 2019. "GHSL Data Package 2019." Technical Report EUR 29788 EN, Publications Office of the European Union.

Freire, Sergio, Kytt MacManus, Martino Pesaresi, Erin Doxsey-Whitfield, and Jane Mills. 2016. "Development of New Open and Free Multi-emporal Global Population Grids at 250 M Resolution."

Proceedings of the AGILE.

Fujita, Masahisa, and Hideaki Ogawa. 1982. "Multiple Equilibria and Structural Transition of Non-monocentric Urban Configurations." *Regional Science and Urban economics* 12 (2): 161–96.

Edward L. Glaeser, Jed Kolko, and Albert Saiz, 2001. "Consumer City." *Journal of Economic Geography,* 1(1): 27–50.

Gollin, Douglas, David Lagakos, and Michael E. Waugh. 2013. "The Agricultural Productivity Gap." *The Quarterly Journal of Economics* 129 (2): 939–93.

Gollin, Douglas, Martina Kirchberger, and David Lagakos. 2017. "In Search of a Spatial Equilibrium in the Developing World." NBER Working Paper 23916.

Gollin, Douglas, Remi Jedwab, and Dietrich Vollrath. 2016. "Urbanization with and without Industrialization." *Journal of Economic Growth* 21: 35–70.

Heblich, Stephan, Stephen J. Redding, and Daniel M. Sturm. 2018. "The Making of the Modern Metropolis: Evidence from London." NBER Working Paper 25047.

Henderson, J. Vernon, Vivian Liu, Cong Peng and Adam Storeygard. 2020. *Demographic and health outcomes by Degree of Urbanisation: Perspectives from a new classification of urban areas*. Brussels: European Commission.

Henderson, J. Vernon, and Sebastian Kriticos. 2018. "The Development of the African System of Cities." *Annual Review of Economics* 10: 287–314.

Henderson, J. Vernon, Sebastian Kriticos, and Dzhamilya Nigmatulina. 2019. "Measuring Urban Economic Density." *Journal of Urban Economics* ISSN 0094–1190.

Ismail, Kareem. 2010. "The Structural Manifestation of the Dutch Disease: The Case of Oil Exporting Countries." IMF Working Paper 10–103.

Kappner, Kalle. 2019. "'Cholera Forcing' and the Urban Water Infrastructure: Lessons from Historical Berlin." EHES Working Paper 0167.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E Waugh, 2018. "The Welfare Effects of Encouraging Rural-Urban Migration." NBER Working Paper 24193.

Lall, Somik Vinay, J. Vernon Henderson, and Anthony J. Venables. 2017. *Africa's Cities: Opening Doors to the World*. The World Bank.

Michaels, Guy, Ferdinand Rauch, and Stephen J. Redding. 2012. "Urbanization and Structural Transformation." *The Quarterly Journal of Economics* 127 (2): 535–86.

Moretti, Enrico. 2010. "Local Labor Markets." In *Handbook of Labor Economics*, edited by Orley Ashenfelter and David Card, 1237–1313. Amsterdam: Elsevier.

Newfarmer Richard S., John Page, and Finn Tarp. 2018. *Industries without Smokestacks: Industrialization in Africa Reconsidered*. Oxford: Oxford University Press.

Quintero, Luis E., and Mark Roberts. 2018. "Explaining Spatial Variations in Productivity: Evidence from Latin America and the Caribbean." Policy Research Working Paper WPS 8560.

Ray, Deepak K., Navin Ramankutty, Nathaniel D. Mueller, Paul C. West, and Jonathan A. Foley. 2012. "Recent Patterns of Crop Yield Growth and Stagnation." *Nature Communications* 3 (1293).

Roback, Jennifer. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6): 1257–78.

Rose, Amy, and Eddie Bright. 2014. "The Landscan Global Population Distribution Project: Current State of the Art and Prospective Innovation." Paper presented at the Annual Meeting of the Population Association of America, Boston, MA, May 1–3.

Sachs, Jeffrey D., and Andrew M. Warner. 2001. "The Curse of Natural Resources." *European Economic Review* 45 (4–6): 827–38.

Sánchez, Perdo. 2010. "Tripling Crop Yields in Tropical Africa." *Nature Geosciences* 3: 299–300.

Schiavina, Marcello, Sergio Freire, Kytt MacManus. 2019. "GHS-POP R2019A - GHS population grid multitemporal (1975–1990–2000–2015)." European Commission, Joint Research Centre, JRC Data Catalogue. https://ghsl.jrc.ec.europa.eu/ghs_pop2019.php (accessed September 1, 2019).

Tombe, Trevor, and Xiaodong Zhu. 2019. "Trade, Migration, and Productivity: A Quantitative Analysis of China." *American Economic Review* 109 (5): 1843–72.

Tsivanidis, Nick. 2019. *The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio.* University of California, Berkeley Haas School of Business.

United Nations. 2018. "2018 Revision of World Urbanization Prospects." World Urbaniztion Prospects 2018. https://population.un.org/wup/ (accessed September 1, 2020).

World Bank. 1981. *World Development Report 1981*. Washington, DC : World Bank Group.

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 174–192

Urban-Rural Gaps in the Developing World: Does Internal Migration Offer Opportunities?

David Lagakos

One prominent feature of virtually every developing country is an enormous divide between rural and urban living standards, measured by income, consumption, or various nonmonetary aspects of life. As a result, much of the inequality within the developing world—home to about half of the planet's nearly 8 billion people—is accounted for by the urban-rural gap.

To illustrate, Table 1 presents a number of comparisons of rural and urban living standards for those residing in Nigeria and India—the most populous countries in Africa and South Asia, respectively—drawing on real outcomes measured from the Demographic and Health Surveys (DHS). In Nigerian and Indian villages, the floor in your home would most likely be made of dirt; in urban areas, floors are most commonly made of wood or stone. About one-half of rural Indians and one-third of rural Nigerians have no toilet facility of any kind—not even a pit latrine or composting toilet—while virtually all urban residents have one, however rudimentary. Fewer than four in ten rural Nigerians can point to a power outlet inside their home, compared with eight out of ten urbanites. Rural Indians similarly lag behind their urban counterparts in electricity connections. In both countries, television ownership rates in cities are about twice as high as in rural areas.

Similar patterns emerge when looking at mortality rates and other health metrics. In both Nigeria and India, you would be just over half as likely to perish before your fifth birthday in a city than in a village. Among adults, a body-mass index

■ *David Lagakos is Associate Professor of Economics, University of California San Diego, La Jolla, California. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His email address is dlagakos@ucsd.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.174.

Table 1
Real Urban and Rural Living Standards in India and Nigeria

	Urban	*Rural*
Percent with finished floors		
India:	70.4	40.3
Nigeria:	88.1	60.8
Percent with toilet facility		
India:	89.5	45.9
Nigeria:	84.6	67.5
Percent with electricity		
India:	97.5	83.2
Nigeria:	82.7	38.9
Percent owning a television		
India:	87.0	53.5
Nigeria:	70.7	30.0
Under-five mortality (per 1,000 births)		
India:	36	59
Nigeria:	86	155
Percent with BMI below 18.5		
India:	15.5	26.8
Nigeria:	9.6	14.4

Note: Compiled from the Demographic and Health Surveys, funded by the US Association for International Development and publicly available at https://dhsprogram.com/. The statistics are calculated in the most recent year available, which is most commonly 2018.

of less than 18.5 is a commonly used indicator of serious malnutrition. For example, someone who is 5' 8" tall (172 centimeters) would have to weigh just 122 pounds (55 kilograms) to have a BMI of 18.5. In both countries your chance of having such a low BMI would be about 50 percent higher in rural areas than in cities.

Of course, well-being depends on a host of other factors other than the few presented here, and some of these are harder to measure. A full accounting of urban-rural differences might also take into consideration the value (positive or negative) placed on the hustle and bustle of urban centers or the security of traditional kinship ties in rural villages. However, the observable differences are so large that it is hard to believe that rural-urban gaps are simply artifacts of inaccurate measurement. The real issue is how to interpret these gaps and whether policy should try to do anything about them.

In this essay, I first set the stage by offering some more systematic evidence on the size and prevalence of the urban-rural gap from a variety of recent data sources. I then discuss whether the urban-rural gap can be explained by sorting; after all, it is clearly the case that those in urban areas tend to have more education, and they are probably selected on other less observable abilities as well. In addition, I review an array of evidence on outcomes of worker migration and find that rural-to-urban migrants do typically obtain higher incomes, which suggests that the pre-migration

situation cannot entirely be due to efficient sorting by skill and education. In addition, while rural-to-urban migrants do experience gains, migration itself does not close the urban-rural gap. Thus, I turn to other potential explanations. I emphasize that a variety of frictions—information, financial, and in land markets—may help to explain the persistence of urban-rural gaps, though much more work is needed here.[1]

Throughout time and space, economic development has been associated with massive population movements from rural to urban areas, and from agriculture to non-agricultural activities. A main reason urban-rural gaps are worthy of study is that it they may be informative about this process of structural transformation. Some writers have pointed to the large urban-rural gaps as suggesting the possibility of massive productivity gains from helping move workers in developing countries out of rural agriculture (Caselli 2005; Restuccia, Yang, and Zhu 2008; Vollrath 2009; McMillan, Rodrik, and Verduzco-Gallo 2014). I offer a more restrained conclusion. Although some rural-to-urban migrants already make gains, many choose not to migrate for a variety of reasons. Some of these reasons may be efficient in the sense that a benevolent social planner wouldn't want them to do any differently. However, others may be held back inefficiently from rural-to-urban migration, and policy may play an important role in reducing the frictions that keep them from the higher living standards that cities would offer them.

Urban-Rural Gaps: Recent Systematic Evidence

Until recently, most of the evidence on urban-rural gaps has come from nominal income or consumption expenditure data by region that has been deflated using spatial price indices. However, price deflation is never quite as straightforward as the breezy descriptions that we (or at least I) offer in undergraduate macroeconomics lectures. The main challenge is that many goods and services are not that easy to compare between cities and rural villages. For example, an urban apartment will carry a higher monthly rent than a straw hut of comparable size in a remote area, but will involve a number of quality differences including access to electricity and running water. This makes apples-to-apples comparisons difficult. Prices for food items are also tricky to compare, because such a large fraction of food in rural areas is home-produced (Deaton and Dupriez 2011).

Ravallion, Chen, and Sangraula (2007) construct arguably the best rural-urban price deflators available for a large set of low-income countries, drawing on spatial

[1] For engaging overviews of urban economics in the developing world, the interested reader might begin with the review articles by Bryan, Glaeser, and Tsivanidis (2019) and Brueckner and Lall (2015) and the book by Glaeser (2017), which helps put today's developing-world cities in a broader historical context. While rural-urban migration will be discussed below, I will focus mostly on how it informs us about the sources of the urban-rural gaps. For a more general review of the literature on rural-urban migration, a useful starting point is the article by Lucas (2015). Similarly, for more on the role of migration and rural-urban gaps in dual economy models in economics, the essay by Gollin (2014) is essential reading.

price data collected by the World Bank. On average, they find that the "same basket," to the extent that such a thing can be measured, costs around 30 percent more in cities than in rural areas. They argue that these price differentials are not nearly enough to offset the higher poverty rates of rural areas that are implied by nominal consumption expenditures. This finding of higher real poverty rates in rural areas is echoed in the work of Ferré, Ferreira, and Lanjouw (2012), among many others, all of whom grapple with similar challenges in making spatial price adjustments.

The pioneering paper by Young (2013) offers an alternative approach to measuring real rural-urban consumption gaps in the developing world that sidesteps the need for spatial price indices. The methodology he develops, while less transparent than approaches based on regional price deflation, allows him to estimate a single proxy for "consumption" per head in rural and urban areas of 67 developing countries using the data from the Demographic and Health Surveys discussed above. The idea is to infer household consumption levels using cross-sectional correlations between educational attainment and the consumption of each "good" measured in the data (Young 2012, 2013). The "goods" are actually 23 real outcomes, including ownership of durable goods (like televisions), housing conditions (like access to electricity), and children's health outcomes. Thus, Young's measure of consumption is broader in some ways than typical national accounts measures of consumption, but it covers a narrower set of goods. Across Young's (2013) set of developing countries, average consumption per household is 4.5 times as high in urban areas as in rural areas. In addition, Young calculates that 40 percent of total inequality is accounted for by the urban-rural gaps themselves. The share of inequality explained by the urban-rural gap is higher in countries with more inequality. For example, in Zimbabwe, one of most unequal countries in Young's data, the urban-rural gap explains close to 70 percent of all inequality.

The urban-rural gaps in consumption and income are related—though clearly not identical—to the gaps in income per head between agricultural and non-agricultural workers. Researchers who focus on agricultural and non-agricultural incomes can make use of national accounts data on value-added by sector, which is widely available. Gollin, Lagakos, and Waugh (2014) draw on sectoral value added data for 151 counties (of all income levels) and construct employment by sector using nationally representative surveys. On average, they find that value added per worker is 3.5 times as high in the non-agricultural sector as in agriculture. This ratio is well above one in all but a handful of countries in their data.

One key concern with sector-level national accounts data is that while the guidelines behind the UN Systems of National Accounts clearly include non-market production of goods (including agricultural goods) as part of value added, in practice such output may be underestimated, as Gollin, Parente, and Rogerson (2004) hypothesize. Indeed, Herrendorf and Schoellman (2015) document this pattern in US agricultural output data. Gollin, Lagakos, and Waugh (2014) seek to address this issue by constructing value added using household income and consumption data for ten countries from the World Bank's Living Standards Measurement Surveys (LSMS), which contain detailed questions about agricultural production both

for own consumption or to be sold in a market. Such measures continue to show large gaps in value added per worker between agricultural and non-agricultural workers. For example, the urban-rural gap in Ghana is 2.2 according to the national income accounts and 2.3 using the household surveys. Cote d'Ivoire has a gap of 4.7 according to the national income accounts and 4 using the household surveys. These multiples are not identical, to be sure, but recalculating based on household survey data continues to leave a large gap.

To circumvent the national accounts data altogether, Herrendorf and Schoellman (2018) focus on wage data by sector, where wages are defined as labor income divided by hours worked. Their data come from nationally representative population censuses compiled by the Integrated Public Use Microdata Series (IPUMS). On average, the simple ratio of non-agricultural average wages to agricultural average is around 1.8 in the 13 countries they study, with a range of 1.5 to 2.7, and gaps of a similar magnitude when splitting non-agriculture into industry and services. These gaps are not as large as some calculated by other methods, but in none of their countries does a simple comparison of average wages yield anything close to parity between the agriculture and non-agricultural sectors.

How do the urban-rural gaps in developing countries compare to those of high-income countries? The limited available evidence points in the direction of larger gaps in developing countries, but this evidence is far from definitive. Drawing on rich regional-level data, Chauvin et al. (2017) document that in China and India, a doubling of population density is associated with an increase in average real wages of around 6 percent. In other words, more urbanized regions offer substantial wage premia relative to more rural areas. In the United States, a doubling of density leads real wages to increase by only about 2 percent, consistent with a smaller urban-rural divide. On the other hand, in Brazil, the wage-density gradient is even smaller at around 1 percent. Looking at agricultural productivity gaps, Gollin, Lagakos, and Waugh (2014) find that the median gap is 4.3 among countries in the bottom quartile of the world income distribution, compared to 1.7 in the top quartile, while the mean gap is 5.6 in the bottom quartile and 2.0 in the top quartile. Thus, these gaps are substantially larger on average in poorer countries. Still, there is a lot of variation within each income quartile, and many high-income countries exhibit very large gaps as well.

A related question is whether urban-rural gaps tend to decrease over time as economies develop and populations urbanize. While many economists have the prior that they do, the historical evidence is thin and some studies have pointed to quite different patterns. For example, Hatton and Williamson (1992) show that the rural-urban gap actually widened in the United States from the late nineteenth century up until right before World War II, and Lundh and Prado (2015) find that the gap in Sweden changed very little over the twentieth century. Hnatkovska and Lahiri (2018) document that, since the 1980s, urban-rural wage gaps have widened in China and decreased substantially in India. A valuable task for future research would be to look systematically at how these gaps have evolved over long time horizons in a broad set of countries.

Sorting and Evidence from Internal Migrants

Urban households differ from rural households in many ways: they perform different jobs, possess different skills, and as one would expect, receive different compensation levels. To what extent can the urban-rural gaps in average incomes be explained by a situation with more productive individuals sorting into cities and less productive types ending up mostly in rural areas? Suppose, for argument's sake, that the answer is 100 percent. Then there would be a sense in which the urban-rural gap would not be policy relevant. In the same way that dentists earn more on average than Uber drivers, urbanites may just be different people from villagers. We should not then be trying to turn villagers into urbanites any more than we should be encouraging Uber drivers to earn their living filling cavities.

In this section, I first discuss some models and quantitative analyses of sorting and then turn to evidence from long-term panel data and experimental studies of internal migration. The theory of sorting makes a strong case that the gains from expanding rural-to-urban migration will be much smaller than a naïve interpretation of the urban-rural gap might suggest because average levels of education and skill are lower in rural areas. However, I will also argue that the reduced-form evidence on sorting can be hard to interpret because these data mix together migrants with many varying motivations for migration—for example, the fact that some migrants are choosing a move toward opportunity while others are making a forced move of necessity—and do not take into account the many varying motivations for non-migration. Overall, sorting alone does not seem to explain the ongoing urban-rural gaps in income and standard of living, though it certainly explains part of it.

A Sorting Approach to Urban-Rural Gaps

At least in cross-sectional data, the signs that there is urban-rural sorting based on observable variables like education are overwhelming. As one example, Gollin, Lagakos, and Waugh (2014) measure average years of schooling among non-agricultural and agricultural workers in 124 countries using census data from IPUMS and household survey data from various other sources. In literally every country, non-agricultural workers have higher average schooling levels, and on average, they have almost twice as many years of schooling than those working in agriculture. The same basic finding shows up in every other cross-sectional comparison of which I am aware, whether looking at non-agriculture versus agriculture or urban versus rural. Taking a historical perspective, Porzio and Santangelo (2019) argue that the rise of schooling globally over the twentieth century was one of the main factors behind the large movement of workers into the non-agricultural sector observed in most countries over this period.

Taking this a step further, Young (2013) looks not only at where workers are currently located, but where they were raised as children. He finds that rural-to-urban migrants have substantially higher education levels than those who were raised in rural areas and stayed there. Urban-to-rural migrants, in contrast, have

much lower education levels than those born in urban areas who remained there. His data cover 170 surveys from a diverse set of developing countries, and thus emphasize that sorting on education is a basic fact of life in the developing world. Theoretical models have been built on this insight; for example, the Lucas (2004) model of rural-urban migration and long-run economic growth begins with the premise that education is valuable only in urban areas.

However, the obvious question is the importance of this sorting in accounting for the urban-rural gaps, and here the literature remains divided to some extent. For example, Gollin, Lagakos, and Waugh (2014) take a basic approach and convert years of schooling into "units of human capital" using an off-the-shelf estimate of 10 percent (Mincerian) return to a year of schooling. Their approach follows the literature on development accounting that does more or less the same thing but in a cross-country setting (for example, see Hsieh and Klenow 2010). The resulting human capital stocks estimated by Gollin, Lagakos, and Waugh (2014) are around 40 percent higher for non-agricultural workers, which explains only a modest fraction of the overall gaps. Country-specific returns and attempts to adjust for schooling quality explain a bit more, but still leave large residual gaps.

Herrendorf and Schoellman (2018) draw on their individual-level wage data to estimate the Mincerian returns to education and experience by sector for their set of 13 countries. They find smaller returns to education among agricultural workers than for workers in other sectors, which squares with the common perception that education is not that useful in farming. Once they add controls, including for education, they find that the raw wage gaps of 1.8 in the median country fall to 1.3. Vollrath (2014) takes a similar approach but by using data from the Living Standards Measurement Surveys for 14 developing countries. His approach includes controls for age and age-squared (to capture experience effects), occupation, and, in one specification, occupation-specific returns to education. His extra controls get the gaps to fall further, though his estimated wages are still lower in agriculture even with the most stringent set of controls.

Of course, economists are fully aware that the measurable variable of education is likely to be highly correlated with a number of unobservable variables, like cognitive abilities or whether a family offers support for education. Furthermore, the importance of sorting is likely to be more important among educated workers than for those performing unskilled manual tasks, among which skill heterogeneity is probably not that substantial. When developing a model for putting education in the context of a broader group of observed and unobservable variables, one sooner or later arrives at the Roy (1951) model of occupational choice. In brief, the Roy model posits that each worker has a vector of "skills"—one for each occupation—rather than a single skill level. Workers then sort into the occupation that yields them the highest income level.

The most difficult challenge with Roy-style models of sorting is that one rarely observes the paths not chosen—that is, the life that rural-urban migrants would have led had they stayed in the village. One approach, first taken in this literature by Lagakos and Waugh (2013), is to draw on more structure to make inferences,

using parameterized versions of the Roy model calibrated to data on wage distributions by sector. The papers of Young (2013) and Herrendorf and Schoellman (2018) introduce richer Roy models that combine sorting on observables, in particular, education as well as unobservable ability. The key theme in these models is that the urban sector is more skill-intensive. Workers are endowed with education, but education is not a skill in and of itself and only increases the probability of becoming skilled. Those becoming skilled stay and work in the skill-intensive urban sector. Those born in the rural area but who become skilled—even with their lower education level—migrate to the city to work in the urban sector. The reverse is true for those not becoming skilled, who find their way to the rural area to work in the unskilled-intensive sector.

The theory is consistent with the clear sorting on education by region in the data, with the more educated primarily locating in cities. Because education and ability are positively correlated, those with higher ability are also primarily located in cities. In the models presented in these papers, the higher average education and ability of workers in cities combine quantitatively to account for nearly all of the urban-rural gap, leaving little or no role for any other explanatory factor.

Compelling as these models may be, there is only so far one can go with sorting stories by looking at cross-sectional data, with one wage outcome per individual (Heckman and Honore 1990). What is more informative, though still no panacea, is to draw on detailed panel data to observe what actually happens to those that move between rural and urban areas.

Panel Data on Internal Migration

Hicks et al. (2017) carry out such an exercise using two long and large household panel surveys: the Indonesian Family Life Survey and the Kenyan Life Panel Survey. What makes these surveys so attractive, besides their length and large sample sizes, is that they make a serious effort to track every respondent that moves between survey waves. Such tracking is not easy: many migrants leave behind little trace of their whereabouts, requiring survey enumerators to do some real detective work.

The punchline of their study is that the gaps between urban and rural areas are far larger than the changes in income and consumption experienced by those moving from rural to urban areas (or from agriculture into a non-agricultural job). In Indonesia, for example, the urban-rural earnings gap is about 1.7 when calculated without any other controls—roughly in line with the cross-sectional estimates described above. But once individual fixed effects are included, the urban coefficient falls nearly to zero. In Kenya, the cross-sectional urban-rural gap in earnings is around 2.4 without any controls, but this falls by two-thirds with individual fixed effects. The results for non-agricultural worker status yield different numbers but the same conclusion: cross-sectional gaps are cut down dramatically once individual fixed effects are added to the regressions.[2]

[2]Hicks et al. (2017) find that cognitive ability scores, measured using Raven's Progressive Matrices, are around 0.3 standard deviations higher in both countries among migrants than for those who remain

More recent evidence from other studies and from countries other than Indonesia and Kenya has tended to find similar patterns; that is, the observed gains from rural-to-urban migration are much smaller than urban-rural cross-sectional gaps. Using detailed wage data from a large survey in Brazil, Alvarez (2020) finds that sectoral movers gain a lot less than one would naively expect given the large cross-sectional wage gap in Brazil. Non-agricultural workers in the cross-section earn a premium of around 62 percent relative to agricultural workers. Once individual fixed effects are included, the premium is a just 9 percent for manufacturing and a paltry 4 percent for services. He concludes that sorting on observables explains close to the entire Brazilian cross-sectional gap.

Lagakos et al. (forthcoming) follow suit by looking at the returns to migration for rural-urban migrants in China, Ghana, Indonesia, Malawi, South Africa, and Tanzania. The surveys they draw on are nationally representative panels and also make substantial efforts to track migrants across space. Like Hicks et al. (2017) and Alvarez (2020), these panel data confirm the substantially smaller returns to rural-urban migration for those choosing to migrate than the cross-sectional gaps. However, their estimated returns are not near-zero after controls for individual fixed effects, and instead, average a substantial 25 percent across their countries. Lagakos et al. (forthcoming) show that their larger average estimated returns come from their different set of countries, not differences in methodology. Their estimated return for Indonesia, also studied by Hicks et al., is the smallest of their six countries.

There are other earlier studies from developing countries that estimate the gains from migrating from urban to rural areas, though these tend to have smaller sample sizes and not to be nationally representative. These earlier studies have generally found substantial returns to migration for those observed to migrate. For example, in a study of 772 rural Indian individuals that were surveyed in 1975 and again in 2005, Dercon, Krishnan, and Krutikova (2013) find consumption per capita was 42 percent higher for those that migrated since the first survey than for those that stayed put. Using panel tracking data from northern Tanzania, Beegle, De Weerdt, and Dercon (2011) find that among 912 households surveyed, those moving out of the community had 36 percent higher consumption levels than those that remained behind after controlling for education, age, and other co-variates. Individuals that moved further away tended to have larger consumption gains.

in rural areas. They cite this, convincingly, as direct evidence that rural-urban migrants are positively selected on ability, in addition to education. While the topic of international migration is beyond the scope of the current essay, it is worth pointing out that the literature on international migration has also found strong evidence of positive selection into migration, both on observables and unobservables. For example, McKenzie, Stillman, and Gibson (2010) provide experimental evidence that the large wage gains for Tongan migrants to New Zealand are largely due to selection on who chooses to apply for a migration lottery. For the much larger set of Mexican migrants to the United States, Chiquiar and Hanson (2005) document strong positive selection on education relative to Mexican non-migrants.

Difficulties in Interpreting Observational Returns to Migration

One might be tempted to conclude that because the gains from migration are substantially smaller than the cross-sectional gaps, there is little scope for policy aimed at encouraging rural-urban migration. Yet this conclusion should be proffered with some skepticism. The "observational returns to migration" estimated from non-experimental panel data require some care to interpret and do not translate as easily as one might think into lessons about the effects of incentivizing others to migrate internally. In general, the concerns relate to the non-random nature of who migrates and to the fact that many people do not actually migrate.

First, worker heterogeneity may extend to migration costs, not just to the migration benefits (as posited in most Roy-style models). Individuals who migrate in equilibrium may be those with relatively low costs and low benefits of migrating, as might be the case for one whose village is close to a major urban center and connected to it by a high quality road. Conversely, those who do not migrate because of high costs—even if they might also experience large potential gains—will never help identify the urban coefficient in a regression that relies only on migrants for identification (Lagakos et al. forthcoming). There is clearly more work to be done to improve our understanding of how migration costs differ across individuals, as opposed to just the returns to migration.

Second, migrating workers who switch sectors may do so for reasons other than choosing the best sector for themselves in a permanent sense. This possibility is consistent with the findings of Pulido and Święcki (2018) who, using the same Indonesian panel data as Hicks et al. (2017), find that around one in five of the movers from the non-agriculture to agriculture sectors describe the shift as "forced" rather than voluntary—for example, when the employer was closed or relocated, or the worker's job was relocated. Those forced to move sectors due to job loss on average witness substantial wage loss.

The underlying problem is that in an observational study, one never really knows what motivates a worker to migrate or not. Once one is "assigned" to migrate using some controlled (or at least well understood) external incentive to migrate, some of the inferences may become clearer. For this reason, many researchers have turned to experimental and quasi-experimental approaches to measuring the returns to internal migration.

Experimental and Quasi-Experimental Returns to Migration

The ideal experiment would be to induce some rural farmers in a developing country to permanently move to urban areas and to observe them and their non-migrant counterparts (plus all of their offspring, while we are at it) for the rest of their lives. But most people aren't likely to move permanently away from their homes in exchange for a modest payment from some experimenting economist. Temporary moves may prove somewhat more feasible to induce, particularly during times when opportunities at home are poor.

In a first-of-its-kind experiment, Bryan, Chowdhury, and Mobarak (2014) tried—quite successfully as it turns out—to induce rural Bangladeshi households to

send migrants to more productive places during the so-called "lean season" between the rice planting and harvest. In the Rangpur region of northern Bangladesh, the lean season brings on a large fall of perhaps one-half in average income, rendering many households so poor as to skip meals. Around one-third of households were already sending out a migrant during the lean season at the time of the experiment, with many going to the urban centers of Chittagong or Dhaka to work as rickshaw drivers, day laborers on construction sites, or some other low-skilled job. Bryan, Chowdhury, and Mobarak (2014)offered households in a randomly selected set of villages an incentive of $11.50 (equal to a few weeks wages) conditional on sending a migrant in the lean season. This modest sum induced a 22 percentage point increase in migration, raising the fraction of households with a migrant from one-third to above one-half. The households with an additional migrant saw consumption rise by a surprising 30 percent per household member. Those in treatment villages were more likely to migrate even three years after the experiment, though only somewhat more than households in the control villages.

A follow-up experiment by Akram, Chowdhury, and Mobarak (2018), carried out in the same region of Bangladesh, offered a richer set of migration incentives and more comprehensive household surveys. The simplest migration incentives, which were about the same size as those in the original experiment, induced a similarly large number of households to send a migrant. The migrants and their families were contacted (pestered, one might say) every week during the lean season with specific questions about the migrants' employment, earnings, and the remittances sent back to their relatives in the villages. In a second treatment arm, a larger fraction of households was offered the incentive in some villages, and this randomly selected second group of villages sent even more migrants. The two treatment arms allow Akram, Chowdhury, and Mobarak (2018) to estimate that rural wages rise by 2 percent for every 10 percent increase in the rural out-migration rate as rural workers become scarcer. In both treatment arms, income and consumption were substantially higher in the treatment villages than in the controls.

Since permanent migration is harder to induce, it is useful to study episodes of forced migration in which individuals in a given area were induced to move out of rural agricultural areas by some strong external force. Perhaps the most relevant study for our discussion is by Sarvimäki, Uusitalo, and Jäntti (2019) who analyze the long-term consequences of when Finland ceded a large portion of its eastern region to the Soviet Union after World War II and had to resettle 430,000 people (11 percent of its population). While Finland is a rich country by any metric today, its GDP per capita was under $5,000 in 1950 (in purchasing power parity terms), and the workforce was mostly agricultural like most developing nations today. A quarter century later in 1971, the groups that were resettled consistently had higher income than comparison groups (like those just on the other side of the border) who had not resettled. The main reason was that being forced to move increased the changes of leaving farming and joining the non-agricultural sector—with its higher wages—by about 50 percent. Interestingly, the children of resettled farmers

also have higher incomes and education than the children of farmers not resettled, pointing to important intergenerational effects of migration.[3]

While all of the evidence presented in this section is specialized in some way, it reinforces the message that the urban-rural gap cannot be solely about sorting. After all, if the rural-urban gap were all about efficient sorting of better workers into urban areas, then an external force inducing people to migrate out of rural agricultural areas should not lead their wages to rise. However, the reasons why these workers were not migrating more often to begin with—and what accounts for the rest of the gap—are not settled. In the next sections, I offer my perspective on two broad classes of possible explanations: compensating differentials of rural life and migration frictions of various sorts. I argue that the latter is the more promising explanation of the two.

Non-Monetary Amenities of Rural Areas

One can easily imagine that the higher wages of developing world cities reflect a premium for lower non-monetary amenities than in rural areas. In fact, this was almost certainly the case in the "killer cities" of the past, which had much higher death rates than the rural hinterlands (Costa and Kahn 2006; Hanlon and Tian 2015; Jedwab and Vollrath 2016). More generally, other non-monetary amenities of rural life may represent the compensating differentials that underly the "spatial equilibrium" assumption common in urban economics in which households are indifferent on average between locations with high wages and fewer amenities and those with lower wages but more amenities (Glaeser and Gottlieb 2009).

In an attempt to shed light on this hypothesis, Gollin, Kirchberger, and Lagakos (2019) analyze spatial data for 20 African countries covering a select number of non-monetary "amenities" related to public goods, pollution, and crime. Theirs is hardly an exhaustive list of all possible amenities but rather, some of the candidates most commonly discussed. They find that public goods are generally much less common in rural areas, including electricity, piped water, and sewage systems (as highlighted in the Nigeria and India comparisons at the start of this article). Indoor air pollution is clearly worse in rural areas because rural households burn solid fuels such as wood for their cooking, which creates a lot of smoke. The World Health Organization (2014) estimates that around four million people die prematurely each year due to burning solid fuels indoors. Perhaps surprisingly, outdoor air pollution is somewhat worse on average in rural areas in Africa. The rural areas in this study tend to be closer to the Sahara Desert where fine particulate matter

[3] In a related study, Nakamura, Sigurdsson, and Steinsson (2019) study the after-effects of a volcanic eruption on a rural fishing community in Iceland which destroyed a random selection of houses. Those under 25 at the time of the eruption who migrated had earnings 83 percent higher than those that stayed behind, some of which happened because the movers completed 3.5 more years of schooling. Bazzi et al. (2016) study an episode of forced migration across islands in Indonesia, in which rural farmers were moved from denser to less dense islands. The overall wage gains were not that large, suggesting that moving workers within the agricultural sector may not be a fruitful way to raise overall productivity.

in the air is highest, while African cities have low levels of manufacturing activity given their level of GDP per capita (Gollin, Jedwab, and Vollrath 2016).

Crime is the one area where African cities appear worse on some metrics than their hinterlands, but the differences are not dramatic. In rural areas, 10 percent of respondents reported that they or a household member were physically attacked in the last year; in urban areas, the rate was 12 percent. Rates of theft are modestly higher in urban areas. When asked about whether they ever felt unsafe in their homes, 37 percent of those in rural areas answered in the affirmative compared to 45 percent in urban areas.

Does evidence on nonmonetary amenities from African cities jibe with the situation in South Asia or other parts of the developing world? More systematic evidence is needed here. In the dimension of air pollution, for example, cities in India have some of the worst air quality in the world and appear much worse than rural areas. In terms of crime patterns, some cities like Bangkok and Manila are thought to have much higher crime rates than their rural hinterlands. Yet in Madagascar, Fafchamps and Moser (2003) find that *rural areas* have higher rates of homicide, burglary, and insecurity than do urban households. Other non-monetary amenities that have not been systematically explored in a developing world context include commuting times and sanitation—these could certainly play some role in explaining some of the urban wage premium. Yet I am skeptical that, taken as a whole, they will explain all that much.

Arguably a more promising version of the amenities story is one in which individuals have idiosyncratic tastes for rural and urban amenities, as in the recent work in the urban economics literature (for example, Kline and Moretti 2014). The idea is that some rural individuals ("the country mice") may optimally choose to remain there even if moving to the city would result in substantial income gains. For example, some rural residents may particularly value living in sparsely populated areas or the bucolic way of life. Such a story can help reconcile the persistence of urban-rural gaps despite income gains for those induced to migrate. The task of separating the idiosyncratic taste shocks from the frictions that hold back migration is a worthwhile—and challenging—job for future research.

Frictions: Information, Financial, and Land Markets

Few economists would dispute the notion that markets in developing countries are full of frictions. A growing body of evidence suggests that some of these frictions may be important factors holding back rural-urban migration and that migration frictions more generally lead to substantially lower aggregate productivity (for example, Bryan and Morten 2018; Tombe and Zhu 2019).[4]

[4]Interestingly, one seemingly obvious type of migration friction—poor road networks—seems to have rather modest effects on internal migration rates and aggregate productivity. At least, this is the conclusion reached by Asher and Novosad (2020); Banerjee, Duflo and Qian (2020); and Morten and Oliveira (2018) in their studies of road building projects in India, China, and Brazil, respectively.

For starters, one wonders how much people in rural parts of the developing world even know about wages and living conditions in distant cities. If most rural residents are not even aware of how much higher wages are in cities, it might help to explain why internal migration rates are not higher. Bryan, Chowdhury, and Mobarak (2014) put this information-frictions theory to the test in one arm of their experimental setting in 16 poor rural villages in northern Bangladesh. Certain households were presented with information on the most common jobs available for seasonal migrants like rickshaw driver, construction worker, or day laborer in four common destinations along with likelihood of finding one of these jobs and average wages. As it turns out, this information treatment had a precise zero effect on migration. In this setting—and remember, around one-third of households were already sending seasonal migrations from this region—it seems natural to assume that households already had sound information about job prospects for seasonal migrants.

However, in a study about migration expectations among rural Kenyans, Baseler (2019) reaches virtually the opposite conclusion. Even though uneducated Kenyans earn twice as much as their rural counterparts, rural workers substantially underestimate the magnitude of this wage gap. To understand why, Baseler (2019) runs an information experiment on a set of villagers where he informs them of the average wages and prices of food in Nairobi and other urban centers, plus the most common jobs for migrants in each potential destination. This simple intervention raises expectations about average urban wages and increases migration to Nairobi, from 20 percent of rural households to 28 percent. Two years later, migration rates were still higher among those getting the information treatment, and migrants reported higher subjective well-being on average.

So why are these Kenyans so poorly informed to begin with? Baseler (2019) theorizes that migrants tend to underreport their earnings in cities to their rural brethren so as not to have to send back too much of their income as remittances. To test his theory, Baseler (2019) runs a second experiment where he spills the beans about all the hidden savings by families with migrants to a random set of others in the villages. This information treatment group responds by updating their expectations about urban wages and their plans to send more migrants in the future. The lesson is that even if out-migration is common in rural areas, the locals still may have imperfect information about wages in the cities. The extent to which information frictions hold back migration in other settings is certainly a topic worth more exploration, especially given how cheap it is to provide information about urban wages and job prospects relative to, say, the cost of providing an additional year of formal education.

Financial frictions of various sorts have long been thought to be important barriers to migration. In particular, financial frictions might bite if migration to cities is a risky enterprise, as emphasized in the classic paper by Harris and Todaro (1970), and potential migrants face borrowing constraints. Bryan et al. (2014) interpret the outcomes of their migration experiments (discussed earlier)

as being about migration risk and borrowing constraints, which keep productive rural people from moving while they do not have a sufficiently large buffer stock of savings to self-insure.

But on further inspection, the explanation that migration risk and credit constraints are what held back migration in this setting doesn't seem quite right.[5] Lagakos, Mobarak, and Waugh (2019) reinterpret the experimental evidence of Bryan et al. (2014) and argue that the experimental data are more consistent with a model in which rural households generally prefer to be in rural areas and only migrate to cities when they are desperate enough to make it worth their while. In the data, it is the households with *lower* assets and consumption levels that are more likely to seasonally migrate, rather than the other way around. Also, when offered cash that is not tied to migration, few households actually decide to migrate in response. Kleemans (2015) finds a similar result in Indonesia, where rural households—particularly those with low asset levels—send more temporary migrants in response to negative rainfall shocks. Lagakos, Mobarak, and Waugh (2019) use their model to simulate the effects of permanently offering migration subsidies and show that the welfare effects from such a policy come largely through offering better insurance to vulnerable rural households.

Munshi and Rosenzweig (2016) and Morten (2019) make the case that a related type of financial friction—the lack of insurance markets—is a constraint on migration that holds down average productivity levels. The idea is that many people are stuck in rural villages because they lack formal insurance, including public safety-net programs, and thus rely on the informal risk-sharing arrangements prevalent in rural communities (Townsend 1994; Udry 1994). In support of this theory, Munshi and Rosenzweig (2016) show that in a set of Indian villages, those that have riskier incomes are less likely to have migrant members. Counterfactual simulations from their model predict that improving insurance markets would lead to substantial reductions in the misallocation of rural workers, many of whom are currently stuck inefficiently in rural areas.

Frictions in land markets in poor countries also may be an important factor holding back rural-urban migration and income levels for many rural households. Unlike in the United States, where real estate is bought and sold readily just about anywhere, few Africans or South Asians hold a title to their land that could allow them to sell it, even in principle. To be sure, land-titling programs are growing, but traditional systems of land rights are still the norm, and markets for land are still nonexistent in most rural areas. In the context of internal migration, the lack of land markets, along with financial frictions discussed above, suggest that it will

[5] Though financial constraints certainly may play an important role in holding back internal migration in other settings, as Cai (2020) shows for China. He randomizes the rollout of a microfinance program in rural Chinese villages and shows that those getting micro-finance loans are much more likely to send migrants seasonally to nearby cities. Migrants experienced increases in earnings of around 36 percent relative to the mean of the control group. Angelucci (2015) and Bazzi (2017) provide evidence that financial constraints hold back *international* migration among very poor households in Mexico and Indonesia, respectively.

be hard for villagers to save up to fund a migration episode, which requires liquid wealth. In addition, those without formal title may be unlikely to migrate for fear of losing their land.

Land market frictions can be potent in holding back migration. A study by de Janvry, Emerick, Gonzalez-Navarro, and Sadoulet (2015) analyzes the rollout of a new land-titling system in Mexico that began in the 1990s, which gave official ownership of land to rural households that had previously farmed the land informally. The authors find that those receiving a title over their land were 28 percent more likely to send migrants. Their interpretation is that the insecurity over land ownership, rather the lack of liquidity, was the main reason migration was not higher before the titling program. In related studies, Chen (2017) and Gottlieb and Grobovsek (2019) argue that land frictions like these, which keep too many workers in agriculture, lead to substantial misallocation of talent. Using a model of structural change calibrated to Ethiopia, Gottlieb and Grobovsek (2018) simulate the effects of improving markets for land and find that it would raise aggregate productivity by 9 percent, as previously misallocated agriculture workers move into more productive non-agricultural activities. Chen (2017) finds an even larger effect in Malawi, which had virtually none of its land titled as recently as 2007.

There are certainly other frictions that hold back migration within developing countries, and research on this important topic is still in its infancy in many ways. Future work should thus continue to help identify and measure the frictions holding back migration of rural workers into more productive opportunities in cities. More analyses that help quantify their importance of specific barriers to migration for development outcomes, like aggregate productivity, would also be most valuable.

■ *For helpful comments on this essay, I thank the JEP editors: Gordon Hanson, Enrico Moretti, Heidi Williams, and especially Timothy Taylor, plus Doug Gollin, Jim Rauch, and Todd Schoellman.*

References

Akram, Agha Ali, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2018. "Effects of Emigration on Rural Labor Markets." Unpublished.

Alvarez, Jorge A. 2020. "The Agricultural Wage Gap: Evidence from Brazilian Micro-data." *American Economic Journal: Macroeconomics* 12 (1): 153–73.

Angelucci, Manuela. 2015. "Migration and Financial Constraints: Evidence from Mexico." *Review of Economics and Statistics* 97 (1): 224–28.

Asher, Sam, and Paul Novosad. 2020. "Rural Roads and Local Economic Development." *American Economic Review* 110 (3): 797–823.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian. 2020. "On the Road: Access to Transportation Infrastructure and Economic Growth in China." *Journal of Development Economics* 145: Article 102442.

Baseler, T. 2019. "Hidden Income and the Perceived Returns to Migration: Experimental Evidence from Kenya." Working Paper, University of Rochester.

Bazzi, Samuel. 2017. "Wealth Heterogeneity and the Income Elasticity of Migration." *American Economic Journal: Applied Economics* 9 (2): 219–55.

Bazzi, Samuel, Arya Gaduh, Alexander D. Rothenberg , and Maisy Wong. 2016. "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia." *American Economic Review* 106 (9): 2658–98.

Beegle, Kathleen, Joachim De Weerdt, and Stefan Dercon. 2011. "Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey." *Review of Economics and Statistics* 93 (3): 1010–33.

Brueckner, Jan K., and Somik V. Lall. 2015. "Cities in Developing Countries: Fueled by Rural-Urban Migration, Lacking in Tenure Security, and Short of Affordable Housing." In *Handbook of Regional and Urban Economics*, edited by Gilles Duranton, J. Vernon Henderson, and William Strange, vol. 5A, 1399–1456. Amsterdam: Elsevier.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak. 2014. "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh." *Econometrica* 82 (5): 1671–1748.

Bryan, Gharad, Edward Glaeser, and Nick Tsivanidis. 2019. "Cities in the Developing World." NBER Working Paper 26390.

Bryan, G. and M. Morten. 2014. "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia." *Journal of Political Economy* 127 (5): 2229–68.

Cai, Shu 2020. "Migration under Liquidity Constraints: Evidence from Randomized Credit Access in China." *Journal of Development Economics* 142: 1–17.

Caselli, Francesco. 2005. "Accounting for Cross-Country Income Differences." In *Handbook of Economic Growth*, edited by Philippe Aghion and Steven N. Durlauf, 679–741. Amsterdam: Elsevier.

Chauvin, Juan Pablo, Edward Glaeser, Yueran Ma, and Kristina Tobio. 2017. "What is Different about Urbanization in Rich and Poor Countries? Cities in Brazil, China, India and the United States." *Journal of Urban Economics* 98: 17–49.

Chen, Chaoran. 2017. "Untitled Land, Occupational Choice, and Agricultural Productivity." *American Economic Journal: Macroeconomics* 9 (4): 91–121.

Chiquiar, Daniel, and Gordon H. Hanson. 2005. "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States." *Journal of Political Economy* 113 (2): 239–81.

Costa, Dora L. and Matthew E. Kahn. 2006. "Public Health and Mortality: What Can We Learn from the Past?" In *Public Policy and the Income Distribution*, edited by Alan J. Auerbach, David Card, and John M. Quigley, 359–98. New York: Russell Sage.

De Janvry, Alain, Kyle Emerick, Marco Gonzalez-Nava Rro , and Elisabeth Sadoulet. 2015. "Delinking Land Rights from Land Use: Certification and Migration in Mexico." *American Economic Review* 105 (10): 3125–49.

Deaton, Angus, and Olivier Dupriez. 2011. "Spatial Price Differences within Large Countries." Unpublished.

Dercon, Stefan, Pramila Krishnan, and Sofya Krutikova. 2013. "Changing Living Standard in Southern Indian Villages 1975–2006: Revisiting the ICRISAT Village Level Studies." *Journal of Development Studies* 49 (12): 1676–93.

Fafchamps, Marcel, and Christine Moser. 2003. "Crime, Isolation, and Law Enforcement." *Journal of African Economies* 12 (4): 625–71.

Ferré, Céline, Francisco H. G. Ferreira, and Peter Lanjouw. 2012. "Is There a Metropolitan Bias? The Relationship between Poverty and City Size in a Selection of Developing Countries." *The World Bank Economic Review* 26 (3): 351–82

Glaeser, Edward. 2011. *Triumph of the City*. New York, New York: Penguin Group.

Glaeser, Edward L., and Joshua D. Gottlieb. 2009. "The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States." *Journal of Economic Literature* 47 (4): 983–1028.

Gollin, Douglas. 2014. "The Lewis Model: A 60-Year Retrospective." *Journal of Economic Perspectives* 28 (3): 71–88.

Gollin, Douglas, Remi Jedwab, and Dietrich Vollrath. 2016. "Urbanization with and without Industrialization." *Journal of Economic Growth* 21: 35–70.

Gollin, Douglas, Martina Kirchberger, and David Lagakos. 2019. "Do Urban Wage Premia Reflect Lower

Amenities? Evidence from Africa." Unpublished.
Gollin, Douglas, David Lagakos, and Michael E. Waugh. 2014. "The Agricultural Productivity Gap." *Quarterly Journal of Economics* 129 (2): 939–93.
Gollin, Douglas, Stephen L. Parente, and Richard Rogerson. 2004. "Farm Work, Home Work and International Productivity Differences." *Review of Economic Dynamics* 7 (4): 827–50.
Gottlieb, Charles, and Jan Grobovšek. 2019. "Communal Land and Agricultural Productivity." *Journal of Development Economics* 138: 135–52.
Hanlon, W. Walker, and Yuan Tian. 2015. "Killer Cities: Past and Present." *American Economic Review: Papers & Proceedings* 105 (5): 570–75.
Harris, John R., and Michael P. Todaro. 1970. "Migration, Unemployment and Development: A Two-Sector Analysis." *American Economic Review* 60 (1): 126–42.
Hatton, Timothy J., and Jeffrey G. Williamson. 1992. "What Explains Wage Gaps between Farm and City? Exploring the Todaro Model with American Evidence, 1980–1941." *Economic Development and Cultural Change* 40 (2): 267–94.
Herrendorf, Berthold, and Todd Schoellman. 2015. "Why Is Measured Productivity So Low in Agriculture?" *Review of Economic Dynamics* 18 (4): 1003–22.
Herrendorf, Berthold, and Todd Schoellman. 2018. "Wages, Human Capital, and Barriers to Structural Transformation." *American Economic Journal: Macroeconomics* 10 (2): 1–23.
Hicks, Joan Hamory, Marieke Kleemans , Nicholas Y. Li, and Edward Miguel. 2017. "Reevaluating Agricultural Productivity Gaps with Longitudinal Microdata." NBER Working Paper 23253.
Hnatkovska, Viktoria, and Amartya Lahiri. 2018. "Urbanization, Structural Transformation and Rural-Urban Disparities in China and India." Unpublished.
Hsieh, Chang-Tai, and Peter J. Klenow. 2010. "Development Accounting." *American Economic Journal: Macroeconomics* 2 (1): 207–23.
Jedwab, Remi, and Vollrath, Dietrich. 2016. "The Urban Mortality Transition and the Rise of Poor Mega-Cities." Unpublished.
Kleemans, Marieke. 2015. "Migration Choice under Risk and Liquidity Constraints." Unpublished.
Kline, Patrick, and Enrico Moretti. 2014. "People, Places, and Public Policy: Some Simple Welfare Economics of Local Economic Development Programs." *Annual Review of Economics* 6: 629–62.
Lagakos, David, Samuel Marshall, Ahmed Mushfiq Mobarak , Michael E. Waugh, and Corey Vernot. Forthcoming. "Migration Costs and Observational Returns to Rural-Urban Migration." *Journal of Monetary Economics*.
Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E. Waugh. 2019. "Welfare Effects of Encouraging Rural-Urban Migration." Unpublished.
Lagakos, David, and Michael E. Waugh. 2013. "Selection, Agriculture, and Cross-Country Productivity Differences." *American Economic Review* 103 (2): 948–80.
Lucas, Robert E. B. 2015. "Internal Migration in Developing Countries: An Overview." KNOWMAD Working Paper 6.
Lucas, Robert E., Jr. 2004. "Life Earnings and Rural-Urban Migration." *Journal of Political Economy* 112 (S1): 29–59.
Lundh, Christer, and Svante Prado. 2015. "Markets and Politics: the Swedish Urban-Rural Wage Gap, 1865–1985." *European Review of Economic History* 19 (1): 67–87.
McKenzie, David, Steven Stillman, and John Gibson. 2010. "How Important Is Selection? Experimental vs. Non-Experimental Measures of the Income Gains from Migration." *Journal of the European Economic Association* 8 (4): 913–45.
McMillan, Margaret, Dani Rodrik, and Íñigo Verduzco-Gallo. 2014. "Globalization, Structural Change, and Productivity Growth, with an Update on Africa." *World Development* 63: 11–32.
Morten, Melanie. 2019. "Temporary Migration and Endogenous Risk Sharing in Village India." *Journal of Political Economy* 127 (1): 1–46.
Morten, Melanie, and Jaqueline Oliveira. 2018. "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City." NBER Working Paper 22158.
Munshi , Kaivan, and Mark Rosenzweig. 2016. "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap." *American Economic Review* 106 (1): 46–98.
Nakamura, Emi, Jósef Sigurdsson, and Jón Steinsson. 2019. "The Gift of Moving: Intergenerational Consequences of a Mobility Shock." Unpublished.
Porzio, Tommaso, and Gabriella Santangelo. 2019. "Does Schooling Cause Structural Transformation? Unpublished.

Pulido, José, and Tomasz Święcki. 2018. "Barriers to Mobility or Sorting? Sources and Aggregate Implications of Income Gaps across Sectors and Locations in Indonesia." Unpublished.

Ravallion, Martin, Shaohua Chen, and Prem Sangraula. 2007. "New Evidence on the Urbanization of Global Poverty." *Population and Development Review* 33(4): 667–701.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu. 2001. "Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis." *Journal of Monetary Economics* 55 (2): 234–50.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu. 2008. "Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis. Journal of Monetary Economics 55 (2): 234–50.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (2): 135–46.

Sarvimäki, Matti, Roope Uusitalo, and Markus Jäntti. 2019. "Habit Formation and the Misallocation of Labor: Evidence from Forced Migrations." Unpublished.

Tombe, Trevor, and Xiaodong Zhu. 2009. "Trade, Migration, and Productivity: A Quantitative Analysis of China." *American Economic Review* 109(5): 1843–72.

Tombe, Trevor, and Xiaodong Zhu. 2019. "Trade, Migration and Productivity: A Quantitative Analysis of China." American Economic Review 109 (5): 1843–72.

Townsend, Robert M. 1994. "Risk and Insurance in Village India." *Econometrica* 62(3): 539–91.

Udry, Christopher. 1994. "Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria." *The Review of Economic Studies* 61 (3): 495–526.

Vollrath, Dietrich. 2009. "How Important are Dual Economy Effects for Aggregate Productivity?" *Journal of Development Economics* 88 (2): 325–34.

Vollrath, Dietrich. 2014. "The Efficiency of Human Capital Allocations in Developing Countries." *Journal of Development Economics* 108: 106–18.

Young, Alwyn. 2012. "The African Growth Miracle." *Journal of Political Economy* 120 (4): 696–739.

Young, Alwyn. 2013. "Inequality, the Urban-Rural Gap, and Migration." *The Quarterly Journal of Economics* 128 (4): 1727–85.

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 193–219

How You Can Work to Increase the Presence and Improve the Experience of Black, Latinx, and Native American People in the Economics Profession

Amanda Bayer, Gary A. Hoover, and Ebonya Washington

The numbers are dismal. Black, Latinx, and Native American people earned just 9.5 percent of economics PhD degrees awarded to US residents in 2017 (National Center for Science and Engineering Statistics 2018). These groups comprise over 30 percent of the population. While the number of Latinx PhD degree earners in economics has increased over recent years, when normalized by their increasing share in the population, the trend is flat. More disheartening, the share of Black economics doctorates has actually fallen since the start of the millennium, while at the same time, the share of Black PhD earners in the fields of science, technology, engineering, and mathematics has risen. There are too few Native Americans earning doctorates to calculate meaningful trends (National Center for Science and Engineering Statistics 2018).

The experiences of minorities in economics are perhaps even more troubling. Of the Black, Latinx, and Native American respondents to the American Economic Association Climate Survey in winter 2018–19, 28 percent report having personally been discriminated against or treated unfairly on the basis of race/ethnicity by someone in the field of economics. Three-fifths of Black and Latinx women, both students and professionals, report experiencing either racial discrimination or gender discrimination or both. These groups are also the most likely to take costly

■ *Amanda Bayer is Professor of Economics, Swarthmore College, Swarthmore, Pennsylvania. Gary A. Hoover is Professor of Economics, University of Oklahoma, Norman, Oklahoma. Ebonya Washington is Professor of Economics, Yale University, New Haven, Connecticut. Their email addresses are abayer1@swarthmore.edu, ghoover@ou.edu, and ebonya.washington@yale.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.193.

action, such as leaving a job, to avoid possible harassment, discrimination, or unfair treatment (Allgood et al. 2019).

The poor numbers and experiences both imply harm to our profession. First, the field of economics is losing out on diverse perspectives and viewpoints. In a variety of settings, racially diverse groups have been shown to outperform homogenous teams (for a review of the literature, see Bayer and Rouse 2016). Second, the underrepresented minority talent we do have is deployed inefficiently, as minorities alter their research, conference participation, and even their workplaces in order to avoid harassment.

You can improve this situation, and this article tells you how. Faced with very little literature to inform the discussion, we collected new survey data. We targeted Black, Latinx, and Native American individuals at various positions in the economics career trajectory as well as those who had been on the trajectory but were no longer. The latter group's views have been much less present in discussions of race and economics. (For example, this group was not reached by the AEA Climate Survey.) We conducted 75 surveys, followed by open-ended interviews with more than half of the participants on the question of what helps and hurts minority group members to succeed in economics. While respondents' date of college graduation and relationship to economics were quite varied, we heard a few consistent themes; namely, that bias, a hostile climate, and a lack of good information and good mentoring discouraged underrepresented minorities from careers in economics.

In response to pressure applied by what was then the Caucus of Black Economists and is now the National Economic Association (NEA), the American Economic Association (AEA) established its Committee on the Status of Minority Groups in the Economics Profession (CSMGEP) over 50 years ago (Simms and Wilson 2020). Soon after, the AEA began its summer training program to prepare undergraduate students from underrepresented backgrounds for PhD programs in economics. In the 1990s, the AEA began its mentoring program, matching underrepresented minority graduate students with mentors. The share of economics PhDs going to Black Americans increased from the mid-1970s until the turn of the millennium, but since that time, as previously noted, it has been on the decline (National Center for Science and Engineering Statistics 2018).

While insufficient dedicated resources inhibit progress, so does insufficient understanding. Very recent developments, including the establishment of the AEA's Code of Professional Conduct, Policy on Harassment and Discrimination, Committee on Equity, Diversity, and Professional Conduct, and Task Force on Best Practices—all since 2018—represent a new awareness of problems on the demand side: the choices of economists at all stages of their (potential) colleagues' careers are reducing the numbers and worsening the experience of Black, Latinx, and Native American individuals in economics.

The reflections and experiences of the respondents —your students, colleagues, and former colleagues—surveyed and interviewed for this paper, point to a variety of actions that you can take to *inform, mentor,* and *welcome.* Some actions, like telling students about programs that help minorities transition from a BA to an economics

PhD, take less effort; others, like starting *new* programs that help minorities make that transition, require much more. Our goal is not that any one person will pursue all of the recommended action steps, but that each economist will take at least one action to improve the representation and experience of Black, Latinx, and Native American individuals in the economics profession.

Survey and Interview Data

We sought to hear from Black, Latinx, and Native American respondents with a great range of experiences in the field of economics.[1] We were aiming for a broad sample, more than a representative sample. In an online Appendix available with this paper at the *Journal of Economic Perspectives* website, we include the survey itself, along with our interview and data coding protocols and procedures. Here, we provide an overview of the process.

We targeted our request for survey participants to the listservs of the American Society of Hispanic Economists (ASHE), the National Economic Association (NEA), and the CSMGEP. Because we also wanted to reach those who had considered an economics career but rejected the option, we also targeted former AEA summer program participants. We asked recipients of our request to take the online survey and/or forward the announcement.

The survey first asked respondents to choose the response that best describes their location on and satisfaction with their economics trajectory. Table 1 shows the possible responses. The key question for analysis in the survey—the *most important question*—asked for open-ended responses to "What is the most important thing we should know about what helps and hurts minorities' progress in an economics career?" Seventy-five eligible people began the survey; 67 answered the *most important question*. We also collected demographic data to determine eligibility for the study, which was based on identification as Black, Latinx, and/or Native American and at least a previous interest in or consideration of economics as a career.

Because we are particularly interested in the factors that lead to a more versus less positive experience in an economics career, we divide the sample into two groups. The first group we term the *nondisrupted* respondents. These respondents' careers are progressing. If they have completed an economics PhD, they are satisfied with their employment. For the purposes of the categorization, we mark the beginning of an economics career as entering a PhD program. Therefore, we also include in the *nondisrupted* group those who never began an economics doctoral

[1]We refer to those identifying as Native American, Indigenous, American Indian, Alaskan Native, and Pacific Islander as Native American, recognizing that there is no consensus on what the most respectful term is. Respondents used the first three, with some respondents using more than one term in the course of the interview. Similarly, the terms Black and Latinx include individuals who identify as African American and, respectively, Hispanic or Latino/Latina. Our focus on these three groups is not a comment on whether those who identify as Asian experience discrimination. The AEA Climate Survey results demonstrate that they do (Allgood et al. 2019).

Table 1
Summary Statistics

	Most important-question sample			*Interview sample*		
	All	*Disrupted*	*Non disrupted*	*All*	*Disrupted*	*Non disrupted*
Race						
Black	47.8	68.8	41.2	50.0	61.5	44.8
Latinx	52.2	31.3	58.8	50.0	38.5	55.2
Native American	7.5	6.3	7.8	9.5	7.7	10.3
Gender						
Women	53.7	68.8	49.0	57.1	76.9	48.3
Undergraduate graduation cohort						
1970s and 1980s	11.9	18.8	9.8	11.9	15.4	10.3
1990s	22.4	37.5	17.6	26.2	53.8	13.8
2000s	31.3	43.8	27.5	31.0	30.8	31.0
2010s	32.8	0.0	43.1	31.0	0.0	44.8
Best describes you						
Considering Economics PhD	14.9	0.0	19.6	11.9	0.0	17.2
Considered Economics PhD but did not attend	16.4	0.0	21.6	14.3	0.0	20.7
In Economics PhD Program	17.9	0.0	23.5	21.4	0.0	31.0
Started Economics PhD but did not complete	7.5	31.3	0.0	9.5	30.8	0.0
Completed Economics PhD, but currently not employed in economics	3.0	12.5	0.0	4.8	15.4	0.0
Completed Economics PhD, started and currently in an academic job, and satisfied	16.4	0.0	21.6	14.3	0.0	20.7
Completed Economics PhD, started and currently in an academic job, and unsatisfied	6.0	25.0	0.0	7.1	23.1	0.0
Completed Economics PhD, started in and left an academic job, currently in a job that uses economic skills, and satisfied	6.0	25.0	0.0	7.1	23.1	0.0
Completed Economics PhD, started in and left an academic job, currently in a job that uses economic skills, and unsatisfied	1.5	6.3	0.0	2.4	7.7	0.0
Completed Economics PhD, started in and went to a non-academic job, currently in a job that uses economic skills, and satisfied	10.4	0.0	13.7	7.1	0.0	10.3
N	67	16	51	42	13	29

Note: Data based on responses to survey.

We define *disrupted* based on the respondents' response to the best-describes-me question on the survey. The respondent was labeled *disrupted* if they answered with one of the following options: started economics PhD but did not complete; completed economics PhD, but currently not employed in economics; completed economics PhD, started and currently in an academic job, and unsatisfied; completed Economics PhD, started in and left an academic job, currently in a job that uses economic skills, and satisfied; completed Economics PhD, started in and left an academic job, currently in a job that uses economic skills, and unsatisfied; completed economics PhD, started in and left an academic job, currently in a job that does not use economic skills; completed Economics PhD, started in a non-academic job, currently in a job that uses economic skills, and unsatisfied; and completed Economics PhD, started in a non-academic job, currently in a job that does not use economic skills.

We omit three responses to the best-describes-me question because no respondents chose those options. They are the following: completed economics PhD, started in and left an academic job, currently in a job that does not use economic skills; completed Economics PhD, started in a non-academic job, currently in a job that uses economic skills, and unsatisfied; and completed Economics PhD, started in a non-academic job, currently in a job that does not use economic skills.

.

program, even if they had considered an economics career and rejected it. The *disrupted* respondents, on the other hand, started on the economics pathway but then left economics, or started on an academic career and then left academic economics, or are unsatisfied in their current position.[2]

Disruption of minority careers is an important feature of the problem to study. Black, Latinx, and Native American students earned nearly 17 percent of bachelor's degrees in economics, but under 10 percent of economics doctorates earned by US citizens and residents in 2018. Subject to selection, survey evidence suggests that the groups make up under 6 percent of US full professors in economics (CSMGEP Annual Report 2019). The shrinkage at each stage suggests the possibility of near-term gains. Relative to their representation in the sample, Black and women respondents are overrepresented among *disrupted* participants, which echoes the AEA Climate Survey results showing that Black respondents of all genders, women of all races, and Black women especially are most likely to have experienced discrimination within economics and to have taken a costly action, such as leaving a job to avoid negative treatment (Bayer 2020).

We followed the survey with open-ended interviews of 42 of the survey respondents. While not common in economics, collecting data through open-ended interviews whose transcripts are then hand-coded by assigning labels to passages of text is a standard form of analysis in other social sciences and is growing in popularity (Miles, Huberman, and Saldaña 2020). We scheduled the interviews for an hour in length, but often with the consent or even encouragement of the interviewee, they lasted longer. Summary statistics for both the *interview* and *most-important-question* samples are in Table 1. Both samples include respondents who graduated from college from the 2010s back to the 1970s, with more robust coverage from the 1990s on. Approximately half of respondents in both samples identify as Black, approximately half as Latinx, and under 10 percent as Native American.[3] (Race categories were not mutually exclusive.)

In targeting survey respondents for our open-ended interviews, we put greater emphasis on following up with *disrupted* respondents. Many are no longer in the economics profession; their opinions are not generally incorporated into discussions on race and economics. Given their original interest in economics, the *disrupted* respondents represent the kind of minority student who might be persuaded to pursue a career in economics, provided we make some changes to the field. In fact, a majority of *disrupted* survey respondents—including all four who began an economics PhD program but did not complete it—would pursue economics if

[2]We categorize those who started and left academia as disrupted even though they remain in economics as a profession, because of the alteration of the career trajectory. Given the desire of respondents to see more professors of color in academia, understanding why respondents left academia is of particular interest. We divide the sample into these two rather coarse groups, instead of more finely, to help maintain confidentiality.

[3]We did not ask respondents for their place of birth or citizenship. Based on responses in interviews, we estimate 69 percent of the interview sample grew up in the United States and 21 percent grew up outside of the United States; we could not determine where 10 percent of the sample were raised.

they had it to do again. In addition to *disrupted* economists, we targeted those with intriguing or unique answers from the *most important question*.

The interview and the *most-important-question* data are complementary. In the interviews, respondents were invited to discuss an unlimited number of factors they felt affect minorities' progression in economics. The interviewers prompted recall of these factors by asking the respondent to walk through their career trajectory. Based on our coding, the top three hindrances mentioned by respondents in interviews were lack of mentoring, lack of good information, and implicit bias. The *disrupted* respondents frequently mentioned those three plus lack of funding, departmental policies/actions, and the hostile climate in economics. *Nondisrupted* respondents cited mentoring and teaching as top hindrances.

The *most important question*, on the other hand, asked respondents to zero in on the most important factor or factors affecting minorities' careers. There is overlap between the hindrances frequently mentioned in the interview and those often mentioned in the responses to the *most important question*, namely around mentoring and implicit bias. However, *disrupted* respondents' views on the *most important question* stood out for their focus on issues about the general climate or their interactions with other individuals in the field, such as implicit bias, elitism, institutional inaction about diversity, lack of understanding/listening on the part of colleagues, bosses, or professors, the field's lack of openness to new questions and methods, and career prospects.

Many respondents expressed feeling unheard in economics: their voices unheard in conversations or seminars and their scholarship unheard in the widest-circulating journals. Because these ideas are their most important contributions to a very significant ongoing conversation on race in economics, we list all responses to the *most important question* in Table 2.

In the remainder of the paper, we synthesize respondents' experiences and reflections, organized not by the hindrances listed in this section, but rather into broad areas of actions that you can take to improve both the numbers of and climate for minorities in economics. The action areas are inform, mentor, and welcome.

Inform

The trajectory to becoming a PhD economist is far from obvious. To take just one example raised by the interviewees, it is not intuitive to undergraduates that an economics major is not sufficient preparation for a doctoral economics program (as highlighted by Sharpe 2017). Because of the dearth of Black, Latinx, and Native American economists and the fact that information spreads more readily within race/ethnicity, students from these minority groups are less likely to receive good information about an economics career. As one graduate student explains:[4]

[4]Throughout the discussion, we describe respondents with minimal demographic information to maintain confidentiality.

Table 2

What Is the Most Important Thing We Should Know about What Helps and Hurts Minorities' Progress in an Economics Career?

Panel A. Disrupted Respondents
Lack of mentorship, also lack of understanding and appreciation about minorities' motivations for entering the field and how that might not align with trendy research paradigms.
Discrimination.
I think access to graduate level classes before entering [a] PhD program would help minorities succeed in completing the PhD course work. I notice a big difference in undergraduate economic courses vs. graduate level. Having some preparation beforehand may have helped in successfully passing preliminary exams.
Helps: Proper mentoring, supportive faculty at the degree granting institutions. Hinders: Discrimination; isolation; racist faculty/institutions.
Networking is important, especially if you are open to exploring non-tenure track academic employment.
Hurts: Access to pre-application process; costs associated to matriculation; biased application selection process. Students not going to certain schools lack access to relationships that would allow minorities to intern or gain jobs in the private sector, government agencies, or NGOs. . . . Haven't seen much overall that helps. Diversity programs are generally targeted towards white women and/or foreign students, when actual actions are taken.
The economics profession is brutal. Colleagues and students can be disrespectful, have implicit biases, and not understand the stress that being a minority economist entails. My senior colleagues also didn't help me with my tenure process, and they didn't help me when stressful situations arose. A better and more efficient network of colleagues and mentors is needed. I don't want to go to meetings and discuss all the issues we know about. Real help and change are needed. More support for minorities is also needed.
I think I suffered from the knowledge that I would not have gotten into my program if it had not been for my skin color and gender. I was clearly unprepared and knew it. I wish I had gotten a masters first, as the majority of people in my program had.
Of course, with seemingly declining opportunities in this digital age where robots are competing with men for jobs, minorities have to make much greater efforts to secure as well as maintain jobs.

(continued)

> [B]ecause there are so many things about advancing in the career that have to do with interactions and with getting the right advice at the right time and when you don't have this knowledge, and when your family doesn't have this knowledge, your friends don't have this knowledge, then it's harder for you to know who to ask first. And even if you know who to ask, it's harder to even approach that person or to go about getting this knowledge.

What Kind of Information?

Respondents argued that more minorities would succeed in economics if they had access to pertinent information at the right time. For example, a first crucial bit of information is how to succeed in college. A recent graduate looks back on her college experience,

> I wish I had been part of a program that would've taught me how college was different from high school. I think that would've made my transition, my first semester 1,000 . . . times easier . . . [S]omebody to tell me, "This is what office

Table 2 (continued)

What Is the Most Important Thing We Should Know about What Helps and Hurts Minorities' Progress in an Economics Career?

Qualifying exams were my road block. I started two well-ranked economics doctoral programs. In the first program, I failed qualifying exams (along with half my cohort). In the second program, I passed two of three qualifying exams and could not continue the program.
The implicit bias within the economics profession concerning minorities contributing to new approaches in economics: New questions, new applied methods, new data collection and survey design to better capture disenfranchised and neglected communities. . . . I was told by senior white males: 'You couldn't have possibly written this article on your own… I was asked by a hiring committee member if I was planning to do 'Latino economics'. . . Deconstructing stereotypes of female and minority research capabilities held by members of the profession requires constant energy and a stainless-steel spine. But if there is no diversity of race/ethnicity, and especially class (working-class members in the profession), nothing will change. The most disappointing time in the academy was training the next generation of minority economics and policy students who would confide: 'Why should I do a PhD in Economics . . . only to face what I see you facing?' . . .*
There is an intolerance [to] differences in appearance and research agendas.
Economics is a wide-ranging field but when it comes down to it, job opportunities in academia are very limited to specific fields and ideas. Hence, we find ourselves in a self-serving and self-re-enforcing loop.
The most important thing that helps minorities is being mentored by other minority economists. The most important thing that hurts minorities is racism. Although many schools are aiming to recruit faculty of color, it is very clear that some places are not open to understanding faculty of color, listening to and supporting their interests, as well as understanding research that is racially motivated.
The economics field is very cutthroat and hostile in general to anyone who does not look like the 'typical' economist. The field has also become too much about prestige and who has the best math skills. Many minorities get into economics because they want to make a difference in their communities, but the way the profession and economic careers are structured, this becomes exceedingly difficult to do.
White [males] have a virtual monopoly over the position of journal editor [not associate editor]. There is an elite monopoly over control of NSF (and similar) funding. Departmental governance and search committees tend to exclude African Americans, assuming there's an African American in the department.

(continued)

> hours are for. It's okay if the only thing you show up to office hours with is . . .'I literally understand nothing. I don't understand. Nothing you said today in class, none of it registered. Can you explain this to me again?'" I feel like that would've really made a difference.

A second example of needed information is what bridge and mentoring programs are available for those interested in an economics PhD. A recent college graduate explains,

> I did not know this A[E]A [Summer] Program existed at all. No one ever mentioned its existence to me. . . . [H]aving this be known to all economists would be very beneficial so that they can then just tell their students . . . about pre-doctoral opportunities, just things of that nature for people who are interested.

Table 2 (continued)

What Is the Most Important Thing We Should Know about What Helps and Hurts Minorities' Progress in an Economics Career?

Panel B. Non- Disrupted Respondents
Not sure. I can only use my experience.
Faculty that inspire students, suggest programs, and look out for them. In undergrad I had a Hispanic professor tell me about the AEA Summer Program and that changed my life in a major way. In graduate school I had a network of people that helped me along, including professors in graduate school.
Feeling isolated likely hurts most. I never felt this way, but this seems to be an important issue for many.
I would say having a mentor helps guide through the process.
Lack of role models; cultural differences perhaps.
Lack of mentors, exposure to internships/careers, knowledge of conferences/networking groups, confidence, and entry level opportunities.
Lack of mentors.
I believe it is really lack of understanding of what to do with the degree. Economics is not [like a] business [degree] where there is a very clear path on how to apply it, so I think that beyond academia, it is not clear. I know that as a minority, many of us did not grow up in an environment where business or careers are necessarily discussed. My parents were blue collar workers. College was a complete unknown other than a degree will help you. It was really feel your way through it.
Access to mentors. I grossly underestimated my need to see someone, anyone doing what I wanted to do who looked like me! Funding. I was unable to further pursue my interest due to being unable to locate funding to pursue and support myself.
I went to an excellent summer program [for] minority students interested in pursuing a PhD. . . . Though I was encouraged to apply for Ivy League schools by the summer program staff . . . I lacked confidence and thought a good compromise was to get my Masters . . . to transition to a PhD program afterwards. I finished my first year successfully. . . . Then, depression set in . . . wasn't sure I had what it took to become an economist. . . . [T]hough my professors were supportive. . . . As a Black woman I did not have any mentors or role models that looked like me. As a result, I don't think I confided in them about what I was feeling and going through . . . I quit altogether. . . . With all that said, I think programs like the summer Econ program for minorities . . . would do even better if they added an ongoing mentoring component after the program to support students in the graduate school application process and journey . . .*
I lacked concrete information on what my career prospects would be–beyond academia. I didn't even know until recently that I could have worked for the Federal Reserve! I think many minorities feel a sense of needing to graduate college and make money, as opposed to spending four additional years to teach.
The academic market for employing minorities seems 'thin' and unwelcoming, as many colleges/universities, of all tier[s], public/private, have not had minority economists, particularly Black economists, on their faculties in over 50 years.

(continued)

What you can do with an economics degree (BA and PhD) is a third example of information that respondents needed. "The clearest path post economics BA from my university seems to be banking and getting information about other paths seems to be like pulling teeth," reads one response to the *most important question.* Many interview participants expressed a similar frustration. They lacked specifics in terms of where economists work. They wanted to learn more about the relationship between an economics degree and public policy work, both in and out of government. One respondent who considered an economics PhD but decided against it looked back more than two decades to her college years. "I lacked concrete information on what my career prospects would be beyond academia. I didn't even know until recently that I could have worked for the Federal Reserve!"

Table 2 (continued)

What Is the Most Important Thing We Should Know about What Helps and Hurts Minorities' Progress in an Economics Career?

Economics as a discipline still has a lot of work to do to critically deconstruct its systemic (both current and foundational) participation in and relationship with structures of oppression. It is very difficult to imagine belonging in a field that (on whole) doesn't do the work to enact structural change, is very resistant to seriously incorporating other social sciences' work and doesn't even really allow much scholarship that names and/or addresses racism directly.
I actually landed a non-academic job after my PhD and made it back into academia after. . . . (It's not an option in the previous list of scenarios) One thing that hurt me was a lack of funding on my first year of the PhD (unlike the rest of my classmates). It also hurt me to come from an unknown undergraduate institution that lacked a strong economics program (the reason I went to this institution was that my parents couldn't afford tuition at the top schools I got accepted into). Finally, I think that my quirks represented an important barrier. I didn't understand how important it was to fit the mold during the job market.
Proper mentorship is the key.
To have some [reference] that it is similar with us and he/she achieved what we desire.
The lack of social capital that comes from being minority or first-gen is [in] my opinion the thing that hurts progress the most. Everything that puts all this information and connections out in the open and makes them inclusive for minorities helps.
This is the most hostile and worst field I ever entered. Faculty are abusive. Nothing that supports minorities in economics. At this point, I am only doing this because I have to. The faculty are outwardly racist.
Mentorship and not belittling people's efforts are very helpful.
Having consideration [of] the different types of interactions that minorities might be used to, given their cultural background is important. Students can be less likely to ask for help or advice but they still need it. Furthermore, assuming that they have the same basic knowledge is also hurtful.
How to move forward on research. Research is hard for everyone but relying on networks for academic success is not an obvious route for people from minority backgrounds, I suspect. It may be easier for people with more privilege, who then have an easier time academically.
That there are high barriers to entry (for example the GRE).
Role of a supportive department and academic mentor who is invested in their success.
Support for publications in top journals. My dissertation advisor contacted the associate editor of a journal and asked him to meet with me to sketch out how my revise-and-resubmit paper should be reframed to meet the objections of the external reviewers. Getting this paper published in a top journal was central to my staying in academia and getting tenure.

(continued)

Respondents also wished they had known more about economic research and what academic economists do outside of the classroom. They wanted salary information too. One interview respondent learned these things when she participated in the AEA Summer Program. "Once I saw that, that . . . doing research, being an academic could be a career and that I could support myself doing that," then a PhD in economics became a real possibility for this participant.

A final example of needed information is, what it takes to be successful as a new PhD. A graduate student expresses this concern,

> People talk about the hidden curriculum or that sort of thing. And it goes really deep because it's not just about how does tenure work? But also, what do specific tenure requirements look like at specific schools? . . . People who have been around in the discipline for long enough know the subtle differences . . .

Table 2 (continued)

What Is the Most Important Thing We Should Know about What Helps and Hurts Minorities' Progress in an Economics Career?

Lack of mentorship during the economics degree. Lack of mentorship upon entering a position.
Hold me to the same standards as white males.
Communication.
Other than structural discrimination [as a hindrance], I think having a mentorship team and peer study groups are critical for successfully graduating with a PhD and making it through the promotion and tenure process.
Funding is an issue if you come from a minority background. Several classmates get extra support from their families in many ways (most classmates come from upper middle-income households in the USA, or high-income households if they are international students), but usually people with minority backgrounds have no external financial support because most of these students come from a working-class background. This is important because extra expenses do occur, minorities without extra financial support have to be RA/TAs . . . during the summer (time to do research is reduced) and if unexpected emergencies occur, they have a larger negative impact among minorities.
Too much [lip] service; We are not monolithic; Lack of cultural knowledge from the majority group; Racism/sexism; Tell us to shut up and be grateful.
Lack of representation in grad school programs from peers and faculty to do research with. Lack of preparation and good mentorship.
Not having minority faculty similar to you in [your] program.
The lack of people with our shared experiences, both in terms of faculty and student-colleagues, hurts minorities progress in economics careers.
I was an average student; I had never taken an economics class before college, so it was all new to me. I persisted in the degree because I was being challenged intellectually, and I enjoyed it. Unfortunately, all the research opportunities were allocated to the top performers in the department. . . . However, an undergraduate email list advertised a position in the economics department; the position did not rely on any grade point average cut[off] score. I met with the professor, also a minority, and he brought me onto the research team…the professor took me under his wing. . . . The support and guidance that a faculty mentor can provide is critical to what helps minorities progress in an econ career. Arbitrary cut[off] scores, on the other hand, can critically hurt minorities' progress. . . . The primary factor that made me want to pursue an Econ PhD was having a meaningful research experience . . .*
Economics is a tool used most effectively to rob African Americans and to enrich the real minority, the 1 percent. When an African American is in the class, the professors recognize that economic theory gets personal. Because of life experience, Black students can see through the facade and this isn't appreciated by many classically trained economists, who do not appreciate having their theories questioned.
I think it is important for minorities who are considering economics to work with minorities who are already working in economics. I went to a summer program [that] encourages minorities to pursue economics but I don't remember being taught by or meeting minorities who had economics PhDs.

(continued)

Disseminate Information

You can illuminate the "hidden" parts of the field of economics. If you are on a college campus, the work can begin even before students arrive. In a field experiment involving 2,710 students across nine US colleges, Bayer, Bhanot, and Lozano (2019) show that emailing information about a diverse array of topics and researchers within economics to incoming women and minority college students increases their likelihood of completing an economics course in the first semester by 3.0 percentage points, nearly 20 percent of the base rate. Such outreach by individual faculty and departments to students from groups not traditionally associated with economics may help offset the information disadvantage and the biased

Table 2 (continued)

What Is the Most Important Thing We Should Know about What Helps and Hurts Minorities' Progress in an Economics Career?

Number of other minority students; Number of advisors interested in topics minority students want to study; The deficit of the doubt, people not interested to talk to you before knowing your credentials.
Lack of information on ways to pursue economics as a career path.
For the most part, all faculty are very thoughtful and encouraging. Early on, however, some faculty would always ask me to think about how my research questions [adds] value to other social scientists. This nudge led me to discard many of my initial motivations for pursuing a Ph.D. But at the time, I did not see this nudge as particularly evil. As I have reflected on this over the years (rising fifth year), I realize that this nudge disproportionately affects minorities. The social science community consists mostly of WHITE MALES. Underrepresented minorities should not be encouraged to cater to either a specific race or gender. I'll reiterate that these statements probably came from a good place. But in retrospect, faculty should avoid hindering potential research agendas because they are not interesting to white males.
It's a very intimidating environment for minorities. As an economist from a Latin American country, we are required to have more years of education to apply for a PhD (most students who apply have a MA in economics). So Latin American students apply to a US program by their mid/late 20s. Thus, as a woman, I find it difficult to commit to a 6-year program at this age, in an environment that is not friendly or flexible for pregnant women or young mothers.
Access to mentorship.
Discrimination; preconceived notions by others about what you should or [should] not specialize in; resentment by non-minorities; preconceived notions of what being a minority means.
Mentorship. . . . It's not just about having figures that look like you that you can look up to as role models but it's just knowing what to expect… I'm Native American and so talking to another professor of economics who is. . . . Native is helpful to better understand . . . problems that would be unique to us . . . thing that hurts minorities progress in economics… a field that values homogeneity . . . if you are a minority interested in economics, you're an outlier . . . having above a 3.8 with a double major in math and Econ is pretty tough. There's no . . . support between these goal posts that make these expectations more realistic . . . wish there was . . . a post-bac . . . for economics. . . . The summer program is great but . . . there needs to be something longer you can do for a year or two with financial support that helps close any gaps that might prevent you from entering into grad school for Econ.*
In the PhD program, the bad behavior of primarily white men in Economics Departments [hinders] minorities' progress. We are treated like no one else. Others don't believe in us, or if they believe in us, they don't believe in us studying certain topics (stay away from race/ethnicity issues). The thing that is most helpful for minorities is professional organizations for us, run by us. These should be supported, well-funded, with wide ability to create opportunities specifically for URM.

(continued)

recommendations of high school counselors (Francis, de Oliveira, and Dimmitt 2019).

Once students are in class, be explicit about your expectations, including appropriate use of office hours, classroom conduct, and class preparation time and what they can expect from you including e-mail use, hours of availability, and the type of help you are willing to give. Let them know what economists actually do and how they can get there through announcements and links on the course website and more seamlessly through course content (which would also speak to respondents' calls for more applied introductory courses discussed later in this paper).[5]

[5]The advice and resources offered on the AEA website (Bayer et al. 2019), including the videos that expose students to the research that economists do (Div.E.Q. 2019) can be helpful in this regard.

Table 2 (continued)

What Is the Most Important Thing We Should Know about What Helps and Hurts Minorities' Progress in an Economics Career?

I currently attend [Redacted institution], and I am an undergraduate student. In my experience, unequal access to academic resources, unfriendly white faculty members, a lack of Black economists who either teach or are advisors all amalgamate to frustrate students. The clearest path post economics BA from my university seems to be banking, and getting information about other paths seems to be like pulling teeth. Even if you can get an understanding of what is necessary, finding ways to conduct interesting research over the summer, pursue a Masters, etc. is daunting to say the least. I am a low-income first-generation college student and a Black woman in a PWI.
Economics department rankings are based on research and [do] not necessarily reflect the knowledge one obtained. However, if you have a degree from a less well-ranked university the disadvantage is fairly big. A standardized test (e.g. GRE) specific for economics would help to remove this bias.
Speaking from my personal life: minorities (I identify myself) have additional hurdles, struggles, and can mask them to blend in. It becomes at times, overwhelming and difficult to pursue academics even when one has the ability. There's simply more to achieving academic goals than the academic support, financial, and environment. I think I may have needed more time to focus, and work out personal obstacles, but couldn't even figure out that much way back then. It's easy to give up. Making the steps smaller, like a road map, but only providing a few small steps at a time would have helped. I did manage a graduate program, filled with minorities, and [took] courses to understand minority struggles so that I could teach minorities, and support them! I did good at that, being able to identify.
Having champions and role models matters and helps. Apathy, unwillingness to engage with issues of diversity, and disdain hurts.
The most important thing to consider is that racism is embedded in the academy and so racial bias affects minorities (especially Black people, specifically Black women) at every stage.
Recognizing structural inequities in the academic system and on top of that, the barriers in a career in academia. I am a first-generation low-income student who came to [Redacted institution] completely unprepared for the level of education. I suffered greatly due to a lack of proper counseling from my school, and lost faith in myself for the next few years. It wasn't until the end of junior year that I began regaining some confidence in myself thanks to some intervention and only then did I know I wanted a career in economics. By this stage, I was unable to generate As in economics classes that professors like to see before helping students. I am STILL looking for pre-doctoral opportunities to do research, but my previous grades and lack of connections are making it difficult.
Representation and support at all stages matter, given Economics' historical preferences towards white men, support for them is systemically built in to allow them to succeed. This is most visible in the advisers and types of research given attention to. A good adviser brings attention to this, provides topics that are of interest to the student, and also shows what the future can look like to a student who has never seen a minority professor before.

Note: Responses edited to protect anonymity. *Indicates a response, originally over 150 words, that was edited for length.

Outside of the classroom, share information actively. Be explicit about expectations with those you supervise as RAs, interns, or PhD advisees. What do you expect their work product to look like? When do you expect it? How and when should they contact you with questions? Provide clear feedback to let supervisees know how they are performing. Make sure that the young people with whom you work have a sense of the breadth of economics in terms of topics studied, career trajectories, and places of employment. Share resources available for succeeding in such a career with those you supervise, whether they ask for the information or not. Students may not know what they do not know. The fact that you speak to the student about graduate school may move that option more squarely into the realm of possibility.

Invite students and junior colleagues to help with your research, to attend conferences, and to join your field. Invite undergraduates to attend seminars. Departments can host brown bag lunches with professors, "so that we can learn about what cool stuff our professors are doing," a recent college graduate suggested. Even better, allow an undergraduate economics club to drive the agenda for a speaker series. Respondents noted that affinity groups, such as groups for minorities in social science, or women in economics, are good tools of information dissemination as well. Departments can encourage and support such groups.

Introduce

You can facilitate connections among researchers and bring minority researchers into your own network. Related to the problem of lack of information is a lack of connection to key networks. If you know of an early-career researcher who would benefit from interacting with another researcher you know, put the two in touch. Connect people online and at conferences. Share opportunities to submit to, present in, and attend conferences. We know these connections are vital to success. A named professor in the sample attributes his strong publication record to being in a network that includes many journal editors. Another PhD says she is no longer in academia because she never made the connections she needed to get her work out there. "In the end, she says "it's who do you want to hang out with, who do you want to be friends with, who do you feel like you can talk with?"

Inform Yourself

Here we provide some basic information about *a few* of the bridge and mentoring programs available for minority economists and economists in training.

The AEA has three programs. In the AEA Summer Program, undergraduates and recent college graduates from underrepresented backgrounds take coursework to prepare them for a PhD in economics. In the AEA Mentoring Program, underrepresented minority graduate students are matched with PhD economists for one-on-one mentoring. In the AEA Summer Economics Fellows Program, graduate students and junior professors are placed within the research communities of government organizations, while allowing the fellows to work on research projects of their own. You can find additional information on all of these programs on the CSMGEP webpage (https://www.aeaweb.org/about-aea/committees/csmgep/programs).

Another major program in this space is the Diversity Initiative for Tenure in Economics (DITE) which supports underrepresented minority scholars moving from untenured to tenured status. There are also several post-baccalaureate programs that allow students to work as research assistants and complete coursework in preparation for a PhD in economics. The PhD Excellence Initiative at New York University (NYU) is one example that targets underrepresented minorities. For a more complete listing, see the AEA's website on Professional Development Initiatives (https://www.aeaweb.org/about-aea/committees/cswep/programs/resources).

To learn about—or post—additional opportunities for minority economists and economists in training, you can subscribe to the CSMGEP listserv and join the ASHE (https://asheweb.org/), which is "concerned with the under-representation of Hispanic Americans in the economics profession" and the NEA (https://www.neaecon.org/), which "promote[s] the professional lives of minorities within the profession," to be part of their listservs and receive up-to-date information on opportunities for minorities in economics. The Association for Economic Research of Indigenous Peoples (https://www.aeripecon.org/), a new professional association founded in February 2019, hosts events and resources "to facilitate intellectual exchange, foster networking and information sharing, encourage and promote teaching and research on topics related to the social and economic development of Indigenous peoples."

Our survey respondents praised bridge and mentoring programs as helpful not only in skill development, but also in terms of information acquisition and network formation. One recent graduate of the AEA Summer Program said: "A lot of people I still keep in contact with. We all, if someone's applying for a job, we'll check resumes. [If I apply to graduate school] I have people I know who can help me, read my application. And just have a support network in general. I think that was a super valuable experience."

While respondents lauded the programs that were available, they saw a great need for more, particularly for those that aid with the transition from a BA to a PhD in economics. Such programs can take many forms, but all should center on the professional development of the participants. One respondent was admitted to a PhD program that allowed her to take undergraduate math classes at no charge for her first year on campus before moving on to the first-year PhD curriculum in her second year. The AEA Summer Program consists of classroom instruction and exposure to current economic research in preparation for a PhD application. Post-baccalaureate programs that allow participants to combine classroom instruction and research-assistant work are popping up on many campuses. These programs can take as few as one student per year and can often garner grant support. If you are interested in starting such a program at your institution, one place to start might be by viewing the joint CSWEP/CSMGEP Panel "Launching a Professional Development Initiative: A Conversation among Mentoring Veterans, Eager Mentors, and Founders of New Mentoring Initiatives" from the 2020 AEA meetings (available at https://www.aeaweb.org/about-aea/committees/cswep/videos/2020/develop).

Mentor

As much as respondents praised bridge programs, there was one suggestion for improvement from a summer program graduate: "I think programs like [the AEA Summer Program] would do even better if they added an ongoing mentoring component after the program to support students in the graduate school application process and journey." In fact, mentoring was the most frequently named solution to

the lack of racial and ethnic diversity in economics, in both the in-depth interview and *most-important-question* samples. As we previously noted, lack of mentoring was a frequently cited problem.

Given the small numbers of minority PhDs in economics, for mentoring to be a meaningful part of the solution, a large share of those mentors will have to be non-minority economists. In our roles on the CSMGEP, we have heard non-minority economists question whether they can effectively mentor minority economists and would-be economists. The answer is yes. More than two-thirds of interview participants reported that a non-minority economist had acted as an effective mentor at some point in their career.

Mentees do not expect all of their mentoring needs to be met by a single mentor. As one graduate student explained,

> I'm sure it's a common thing for people of color . . . and maybe for women, but you have to cobble together your team of mentors or advisors and have them each serve their own purpose. . . . [M]y advisor is not going to be the person that I listen to about race and ethnicity in research. . . . I think as much as she's read about that, that's not her expertise, and that's not where her political investments really line up with mine. She's good at helping me navigate the program and navigate the job market. And my mentor through the AEA mentoring program [Latinx like the speaker] . . . in terms of political investments and in terms of inspired research, I talk to him and I get a million ideas. . . . And I feel so jazzed about changing the discipline talking to him.

A respondent who started but did not finish a PhD program summed up the situation succinctly. "I think there has to be an effort by non-minority scholars to both understand that there is a pipeline problem . . . and be willing to mentor."

The key phrase is "an effort by non-minority scholars." You will need to do more than wait for a minority student from an underrepresented group to knock on your office door and ask for mentoring. As one student explains,

> First, undergraduate students need to see that there are people like them in the field. But you can't have that unless you have people in the field. So, it's a bit of a vicious cycle. But I think undergraduates also need people to tell them that if you are a minority, you can also come in. So even though you don't look like me, you're also smart or you're also talented or you're also driven enough to do this. So, I think [even] when there's no diversity in the cohort or [among the] professors, they can still actively look for diverse students. And I say actively . . . because I think one of the mistakes a lot of people make is assuming that, oh, the smart and the driven students are the ones that come to meet me [and] are the ones who asked me about graduate school. There are many smart and driven students that just don't dare or don't have maybe the language skills or don't have the information . . .

> [S]ometimes you just don't know. And I think the role that a professor can play, or a mentor can play is really big.

The mentoring relationship may last months or years, may be declared formally or arise more organically. (Please do not let the possibility of an organic mentorship arising prevent you from reaching out to students, from participating in formal mentoring programs, or from even starting a mentoring program.) Specifically, what are students and assistant professors of color looking for a mentor to do? First, provide information and connect the mentee to resources and networks as described earlier. Respondents would like to meet regularly with their mentors. Respondents would also like mentors to listen and help the mentee to become the economist that the mentee wants to be, not a reflection of the mentor. Respondents are also looking for mentors who are proactive. An academic explained:

> You asked me in terms of diversifying the profession and also mentoring and the things I talked about, sponsorship and mentoring, making sure that before the student [starts] failing classes, intervene and make sure, check in with them. How are you doing? What do you need to get through to the classes? . . . What do you need to get through this dissertation? Would you like to coauthor this paper? When they're faculty members, would you like to coauthor these papers? Would you like to go in on this NSF grant? Those are the kinds of things that are needed to retain [underrepresented minorities in economics].

Finally, mentors should be encouraging. Be your mentee's champion. Too many respondents were discouraged from pursuing economics. One was told she would never attain the mathematical proficiency needed. Another was advised that because he was unable to get into a top ten program that it wasn't worth attending economics graduate school, which is a manifestation of the elitism that many interviewees perceived as a major hindrance to their trajectory in economics. A respondent who went as far as an MA in economics was asked about what factors might have led to his earning a PhD, and responded, "I don't know. I guess maybe just being told that you can actually be an economist."

Good mentoring was pivotal to many respondents' decisions of whether to major, work, or get a PhD in economics. One AEA Summer Program graduate believes that had she had a mentor to help her, she would have applied for economics PhD programs when she returned to her home campus that fall. She never did apply, but says if she had to do it again, she would because she got a "thrill" from doing math and economics. More than one respondent credits good mentoring with their being in an economics PhD program today. A student explained the value of mentoring as follows,

> I wouldn't have known from anyone that a PhD is what I needed if I hadn't had someone saying, 'Hey, this is what's going on. This is what [you] probably

> need. I think you can do it.' If we don't have people going out and speaking and encouraging and mentoring and being in relation with us, then what, you know?

For more on the good, the bad, and the ugly of mentoring, see Bogan (2019) as well as Cook (2019) and the *CSWEP News* issue on mentoring underrepresented minority women edited by Mora (2019).

Welcome

Broaden the Introductory Course

From the first contact that students have with economics, the field is off-putting, respondents tell us. Some believe that the introductory courses are designed to "weed out" students from the major. Whether professors are doing this deliberately to decrease numbers or not, respondents point to theories and formulas devoid of applications as uninviting. The impact of these uninviting courses is not equal across student demographics. Bayer et al. (2020) find that minority (and women) students in introductory economics classes report significantly lower measures of relevance, belonging, and growth mindset; for example, they are less likely to agree that their professor uses relatable examples, to report feeling comfortable asking questions in class, to believe that people like them can become economists, and to believe that they could learn the material.

Respondents recommend that introductory courses be more applied and that they include examples that are relevant to students from all backgrounds. In fact, nearly 75 percent of interview respondents cited an interest in public policy in explaining what first attracted them to economics. Says a recent undergraduate,

> Being able to connect what's happened in the classroom with what a given student's lived experience or question is, is extremely useful. . . . I personally know a lot of . . . people of color, who, I think personally, if economics was much more accessible, they would probably be economists because they're interested in questions of, how do we fix the gender gap? How do we fix the racial disparities in education and wages? These are economics questions.

A student who is now in an economics PhD program was hooked by being invited to critique the textbook models,

> Every time in class [the teacher] would say, "I'm going to show you this model[and] I want you to know that these are all of these assumptions baked into this model. . . . Every day you should ask yourself whether those assumptions are really true. . . ." And so, that got me thinking. . . . And I started realizing that I actually wasn't that bad at math. I got to this point where I was like, oh wait, no, I'm proficient, I can probably do this.

Another discouraging factor in one's early years in economics is that, according to several respondents, professors give the impression that they are only interested in the top students. The same recent undergraduate advocating for more applied work above explained, "It's like an ID card. Show your ID, show your A, 'Okay, you get a letter. You get my attention.'" A respondent who graduated from college some time ago talked about how disheartening it was that he could not get a research assistant position on campus because all the positions had requirements for a minimum grade point average. Given that students come into college (and graduate school) with different prior academic experiences, have different rates and manners of assimilating information, and that there are types of intelligence invaluable to research that may not show up on an introductory microeconomics final, we do ourselves (and of course the students) a great disservice by eliminating them from consideration based solely on grades.

Some students will come into an economics course (at the undergraduate or graduate level) with a strong network. Others will know no one and will therefore be at a disadvantage in completing coursework and studying for exams. Level the playing field by formalizing processes and taking across-the-board steps: share resources like copies of past exams universally, set up study groups for all students and encourage students to work collectively, and assign and guide graduate student advising.

Students are receiving implicit and explicit messages about the identity of who belongs in the field. Counter those messages.[6] Be upfront with students about the economics profession's need to be more diverse and the messages of exclusion communicated by materials that omit or diminish the experience of minorities. Div.E.Q. at DiversifyingEcon.org (Bayer 2011) provides strategies for managing diverse classrooms.

Call Out Bad Behavior

The economics profession does not become more welcoming in graduate school and beyond. In fact, interview and survey respondents deemed it as "hostile," "cutthroat," and as previously stated, "elitist." One associate professor describes her department's treatment of graduate students as essentially their saying,

> "We're going to pit you all against each other. We're only going to support the top students. This is a fight to the death.". . . The type of environment that's in economics. I'm just going to be honest with you, I don't think that a student

[6]A number of steps can serve to increase the sense of belonging for minority students, including offering diverse instructors (Fairlie, Hoffmann, and Oreopoulos 2014), teaching assistants (Lusher, Campbell, and Carrell 2018), authors (Bayer, Bhanot, and Lozano 2019; https://jlsumner.shinyapps.io/syllabustool/), and speakers (https://econspeakerdiversity.shinyapps.io/EconSpeakerDiversity/). In addition, successful interventions help students see that adversity is normal rather than an indictment of their belonging, reinforce beliefs that success is attainable and worthwhile, and build community. For example, an intervention designed to buttress Black students' sense of belonging improved academic achievement, as well as self-reported health and well-being, and reduced the reported number of doctor visits for three years after the intervention (Walton and Cohen 2011).

> who is an underrepresented student does well in that environment. I think that we tend to do well in a more supportive environment where we don't feel so isolated.

And about the profession in general, she continued, "The economics profession is brutal. Colleagues and students can be disrespectful, have implicit biases, and not understand the stress that being a minority economist entails."

There is a long history of economists from minority groups being pushed out, neglected, and undervalued. Examples include the experiences of Sadie Tanner Mossell Alexander as the first African-American woman to receive a PhD in economics in 1921 (Malveaux 1991; Banks 2005) as well as the history of the National Economic Association (Simms and Wilson 2020) and the results of the recent AEA climate survey (Allgood et al. 2019).

To create a more welcoming climate in economics for minority scholars, you can call out unacceptable behavior— racism, sexism, harassment of all types—when you see it and when it is reported to you. Your workplace (and our profession) needs clear policies and consequences for this behavior. The AEA Best Practice website (https://www.aeaweb.org/resources/best-practices) provides guidelines for developing such policies.

More often than explicitly racist behavior, respondents had interactions that suggested more subtly expressed racial bias. The associate professor continues,

> [My white] colleagues or even other administrators, they don't know that you have to deal with different layers. . . . I can say something, exactly the same as my white male colleague, but when I say it, I'm being a witch, or I'm mean, but they don't deal with that. . . . So, I'm usually the first Black economist [my students have] ever seen. And that in and of itself has its own issues, like are you qualified? Are you competent? That's what I mean is [my colleagues] don't have to deal with [that] . . . and then when you do get eval[uation]s, . . . what [my chair] noticed was that when it comes to minorities, [the students] veer off [into] personality. He's like, "I've never seen that with eval[uation]s of the white colleagues.". . . They might like my class [but] they'd be like, "But I don't like." And it's usually something really personal or how I dress. And he's like, "I never see that with other colleagues unless they are . . . underrepresented minorities."

Other examples of this sort of subtle bias include differential treatment by colleagues, disrespectful interactions with supervisees or students, receiving more challenges and more interruptions from seminar audiences, and having one's right to be in a certain job or location challenged. To make the economics environment more welcoming, you can raise awareness of the more subtle, but pervasive, biases too. Question whether evaluations, made by students or by colleagues, might be biased. Indeed, survey responses like ours along with the findings of large-scale studies led the AEA to recommend "Do not rely exclusively, or even primarily, on

student evaluations of teaching to inform tenure and promotion decisions" (Bayer et al. 2019). Reduce the influence of remaining biases by standardizing processes such as job searches. Again, the AEA Best Practices website offers details (Bayer et al. 2019).

Listen

When asked what would improve diversity in the economics profession, an economist at a policy organization answered,

> I think people understanding that we are unique, and we all have different experiences. People being open to people talking about their experience and . . . actually hearing them. [I am] not saying that you have to agree with them. You don't have to, that's fine. But really listening to what people are saying that are from different backgrounds and saying like "Wow, I've never experienced that in my entire life but maybe that could be the case. And how can we talk about this?" And being more open about issues that minorities have.

Diversifying the profession means not just diversifying the hue of the skin of the people who do economics, but also diversifying the approaches, questions, experiences, and goals of economists. Respondents asked to be heard on these differences, which shape both the substance of their work and their workplace experiences with implicit and explicit bias. Respondents also wanted to be heard by advisors whose help they would like in reaching research and career goals that are *distinct* from those of their advisors. For example, one PhD economist who is satisfied with her career in industry said, "If you're in graduate school and you know you don't want to be in academics, then there should be . . . someone telling you that's okay; you don't have to be an academic. You can go to industry."

Respondents want colleagues and department chairs to hear that service can be different for minority scholars, not only in terms of quantity of committees, but also in terms of intensity of the work. For example, minority faculty members frequently report more than their fair share of advisees, at least de facto, as minority students are eager to work with a minority advisor. Finally, respondents want to be heard in seminars and in other discussions of research including when they raise critiques through a racial lens, "Instead of being met with . . . not her again," said a PhD student. The point is to listen actively to each individual's particular concern, which may be quite distinct from yours, and then help to address it.[7]

Broaden the "Legitimate" Topics

More difficult than simply trying to raise certain ideas, respondents say, is trying to pursue them as research topics.

[7] Participants in an AEA Panel on "How Can Economics Solve Its Race Problem?" offer other examples of experiences as minority economists and the responses they find helpful. See https://www.aeaweb.org/webcasts/2020/how-can-economics-solve-its-race-problem.

> I think there's a problem with this whole notion of . . . bringing new questions and new ways of approaching very established and old issues in economics. People like to support, especially scholars of color, if they're . . . echoing the mainstream and it's harder and tougher when you're not doing that.

Respondents struggled to get the economics community to engage with work that was viewed as interdisciplinary in nature, was outside of the neoclassical paradigm, or that challenged economic dogma, among other topics. This played out for a current graduate student as follows,

> A lot of the research questions I had . . . as a first-year graduate student were kind of particular to my upbringing and the things that I experienced growing up and as a young adult. Early on, however, some faculty would always ask me to think about . . . my research questions' value to other social scientists. This nudge led me to discard many of my initial motivations for pursuing a PhD, but at the time, I did not see this nudge as particularly evil. As I have reflected on this over the years, I realize that this nudge disproportionately affects minorities. The social science community consists mostly of white males. . . . I'll reiterate that these statements probably came from a good place. But in retrospect, faculty should avoid hindering potential research agendas because they are not interesting to white males.

Different backgrounds and lived experiences of course can lead to different research interests and insights (for example, Bayer and Rouse 2016; May, McGarvey, and Whaples 2014; Malmendier, Nagel, and Yan 2017). Respondents report that advisors and mentors particularly discourage graduate students and early stage minority researchers away from topics related to race or other aspects of their identity, which are the topics that in many cases drew the young researchers to economics in the first place. There is a perception that Black scholars studying Black people or Latinx scholars studying Latinx people or Native American scholars studying Native American people may be biased or taken less seriously as scholars. (Of course, this critique is never made of white men scholars studying white men.) Sadly, this double standard has not changed across the years. A full professor who earned her PhD several decades ago says that because of a nudge away from identity, her early work was "totally sexless and ethnicness-less" and it was 20 years before she began studying a topic that she enjoys and has been productive in, a topic related to her background. Many underrepresented minorities are drawn to economics research because they find the existing research to be problematic or lacking in some fashion.[8]

[8] For example, a frequent critique of the literature on the economics of race is that it relies on a "deficit" model of the behavior in communities of color, rather than racism or structural racism, as explanations for differences in outcomes across race. For discussion, see the Darity et al. (2015) introduction to a special issue of the *Review of Black Political Economy* devoted to stratification economics and the open letter to economists from Bill Spriggs (https://www.minneapolisfed.org/~/media/assets/people/william-spriggs/spriggs-letter_0609_b.pdf).

We imagine that advisors and mentors suggest against certain topics because they believe that conference organizers, journal editors, and hiring and tenure committees will not be appreciative of them. Thus, the gatekeepers in the economics profession need to take a broader view of legitimate economics research.

Engage, Admit, Hire, Promote

Respondents in our survey were unequivocal that having more peers, mentors, and role models of color would improve the economics experience for minorities. One graduate student said it would be "revolutionary." Many underrepresented-minority students struggle with being the only "one" of their race, class, and/or academic background. Some students find it hard to connect across race. More frequently, respondents yearn to be in dialogue with other minority economists. One respondent attended the AEA Summer Program and yet does not remember having a single instructor of color. Another expressed surprise: "Even though I went to college in [name of city], which has a very large minority population, I didn't know any economists who were minorities on a basis where I could shoot ideas off of them or ask them what their story was, or how they got to where they needed to be." Four Black women in the interview sample attribute their failure to go on to get their PhDs at least in part to the lack of mentors who look like them.

Role models can be on syllabi, not just in the classroom. You can increase exposure to the work of underrepresented-minority economists; as one example, social media can be a platform to lift up research by minorities. Diversify the literature you cite on syllabi and in research. Consider racial diversity in the seminars and conferences you organize (and as a tenured respondent to our survey added, not just on the race-related panels). A great new resource is the Diversifying Economics Seminars – Speakers List (https://econspeakerdiversity.shinyapps.io/EconSpeakerDiversity/), which includes economists who identify as underrepresented minorities and/or women and/or LGBTQ+, along with their area of research and contact information.

If you are a journal editor, diversify who you publish. If you are not receiving enough submissions by underrepresented minority economists, reach out and solicit them. Invite minorities onto the editorial team. Respondents pointed to the lack of diversity in terms of both race and institutional affiliation of editors as contributing to the lack of diversity of authors and topics in economics' leading journals. A tenured professor, unsatisfied in her position, put it bluntly: "We need to quit being elitist. If you look at who are the editors of a lot of these journals, it's the same schools. So . . . if econ really wants to increase diversity, we have got to get editors who didn't all go to Harvard, Stanford, Yale, Princeton . . . . " Of course, the more you mention, cite, invite, and publish the work of minority economists, the more likely it becomes that those minority economists will be able to stay in the profession. One way to increase your exposure to the intellectual contributions of minority economists is by

attending sessions sponsored by CSMGEP, NEA, and ASHE at annual and regional meetings.[9]

In the area of admissions and hiring, one cannot rationally expect to obtain a different outcome with the same behavior. If you want to increase diversity, then existing admissions and recruiting practices need to change. We do not know what the final process should look like, nor do we assume it will be the same for every department. But items that should be on the examination table include: First, recruiting from a wide variety of schools including historically black colleges and universities and other minority-serving institutions, and dropping the elitist view of favoring applications from only certain schools. Second, communicating by words and actions that students of all kinds are welcome in the economics profession. Third, discontinuing reliance on the Graduate Record Exam (GRE), an exam on which women and underrepresented minorities score lower on average. The Educational Testing Service (2018), which administers the GRE, warns against using a test cutoff as the sole factor in admission denials. The exam has been dropped completely by top PhD programs in several of the sciences in favor of more holistic evaluations of applicant potential (Langin 2019). Fourth, developing a more holistic (beyond the scores and grades) picture of an applicant's potential research ability. Physicists and others have identified graduate admissions criteria that keep longstanding inequalities in place and have developed methods for recognizing and selecting for unrealized potential in students during the admissions process (Posselt 2016; Stassun et al. 2011).

Hiring and promotion practices need to change, too. Explains the same tenured professor who spoke about economics journal editors,

> And again, it comes from having people who are more open when it comes to reading applications, who [don't] only want an applicant who looks like A. Who are more open to "Hey, this person might not [look] like A, but they're doing some really interesting research. Maybe we should give them a position here. Even though they might not be in our network of friends."

Here are some suggestions to begin a discussion in each department. First, cast your net widely. For junior positions, that means advertising broadly and giving full consideration to applicants from all schools. For senior positions, this means creating a census of potential applicants and not simply considering those economists who come to mind. Second, when we picture the ideal hire, we are overly influenced by what we have seen in the past—the phenotype, the academic record, and the research (Tverksy and Kahneman 1973). Strive to counter this tendency—that is, to work with a broad definition of the questions, approaches, and experience profiles that your department needs. Third, prioritize diversity (of experience and of thought) in searches, or else this concern is likely to fall to last

[9] See the most recent issue of their joint newsletter, *The Minority Report*, for a list of recent sessions and other information at https://www.aeaweb.org/about-aea/committees/csmgep/minority-report.

place. Fourth, structure recruiting and evaluation processes and standardize interviews so that unconscious biases do not create unequal opportunities for candidates to perform and impress. Fifth, at promotion time (or preferably before), note that diverse people will have diverse experiences and recognize and account for these diverse contributions (like an advising overload) to your department. Sixth, hire for the work and not for the phenotype. Respondents loath tokenism. To help get you started on this work, see the Commission on Ethnic Minority Recruitment, Retention, and Training in Psychology (2019) and the AEA Best Practices resource (Bayer et al. 2019).

We will add a word about quality, a concept that is often used to end conversations about diversifying schools and workplaces. A respondent reported on a faculty meeting, "We were talking about what can we do to recruit, especially more Black professors. And then one of my colleagues was like, 'Well, we don't want to lower our standards.'" Darity (2010) discusses the role of "the fetishization of 'merit' as a rationalization for discriminatory outcomes." It is a given that departments do not want to lower standards. That this issue is raised (repeatedly) after someone mentions diversity and hiring is the sort of microaggression that can make minorities feel outraged and unwelcome. It is also wrong. For departments of economics, the existing biases in current admissions and recruiting practices are the equivalent of leaving money on the table. As evidence, note that prominent companies around the world are working to increase diversity (for ten examples, see https://www.socialtalent.com/blog/recruitment/10-companies-around-the-world-that-are-embracing-diversity).

Conclusion

Economics has a diversity problem. The numbers of Black, Latinx, and Native American economists are low, as is their relative level of satisfaction in the economics profession. Based on surveys and interviews of underrepresented minority economists, both those who are on and those who have exited an economics career trajectory, we have offered ways in which you can help to counter this problem, grouped under the action areas of inform, mentor, and welcome.

As economists, we recognize that individuals respond to incentives. Research, teaching, and service are all incentivized, although admittedly to varying degrees. Economists put effort into these activities. The work gets done. But the meager incentives that exist for increasing racial and ethnic diversity across economics have proven insufficient to move the needle. Issuing diversity statements is far from enough. As one PhD economist who is no longer searching for an academic job told us, "I think that there needs to be intensified pressure on economics departments, in particular, to hire people of color. Because until there is pressure, I don't think people are going to change who they're seeking." Forceful incentives have yet to be applied around diversity issues. It is logical to conclude that diversity must not truly be a priority in the economics profession.

■ *We thank interviewers Erika Jackson, Britney Moreira, and Trenel Francis; coders Avery McKenna and Kayla Preito-Hodge; those who assisted us with qualitative methods LaRue Allen, Elizabeth Ananat, Desmond Ang, Jiwon Choi, Sarah Jacobson, Joanna Lahey, Jenny Shen, Lisa Suzuki, and Cirecie A. West-Olatunji; and Trevon Logan and editors Gordon Hanson, Enrico Moretti, Timothy Taylor, and Heidi Williams for helpful comments. A special thank you goes to Arkey Barnett who provided stupendous research assistance.*

References

Allgood, Sam, Lee Badgett, Amanda Bayer, Marianne Bertrand, Sandra E. Black, Nick Bloom, and Lisa D. Cook. 2019. *AEA Professional Climate Survey: Final Report* Nashville, TN: AEA.

Banks, Nina. 2005. "Black Women and Racial Advancement: The Economics of Sadie Tanner Mossell Alexander." *Review of Black Political Economy* (33) 1: 9–24.

Bayer, Amanda, ed. 2011. *Diversifying Economic Quality: A Wiki for Instructors and Departments.* Nashville, TN: AEA CSMGEP.

Bayer, Amanda. 2020. "(Professional) Climate Change." *The Minority Report* 12 (Winter): 16–19.

Bayer, Amanda, and Cecilia Elena Rouse. 2016. "Diversity in the Economics Profession: A New Attack on an Old Problem." *Journal of Economic Perspectives* 30 (4): 221–42.

Bayer, Amanda, Şebnem Kalemli-Özcan, Rohini Pande, Cecilia Elena Rouse, Anthony A. Smith Jr., Juan Carlos Suárez Serrato, and David W. Wilcox. 2019. "Best Practices for Economists: Building a More Diverse, Inclusive, and Productive Profession." *American Economic Association*. https://www.aeaweb.org/resources/best-practices (accessed January 15, 2020).

Bayer, Amanda, Syon P. Bhanot, and Fernando Lozano. 2019. "Does Simple Information Provision Lead to More Diverse Classrooms? Evidence from a Field Experiment on Undergraduate Economics." *AEA Papers and Proceedings* 109: 110–14.

Bayer, Amanda, Syon P. Bhanot, Erin Bronchetti. and Stephen O'Connell. 2020. "Diagnosing the Learning Environment for Diverse Students in Introductory Economics: An Analysis of Relevance, Belonging, and Growth Mindsets." *AEA Papers and Proceedings* 110: 294–98.

Becker, Charles M., Cecilia Elena Rouse, and Mingyu Chen. 2016. "Can a Summer Make a Difference? The Impact of the American Economic Association Summer Program on Minority Student Outcomes." *Economics of Education Review* 53: 46–71. https://doi.org/10.1016/j.econedurev.2016.03.009.

Bogan, Vicki L. 2019. "Academic Mentoring Relationships: The Good, the Bad, and the Ugly." *AEA Minority Report* 11 (Winter): 18–20.

Commission on Ethnic Minority Recruitment, Retention, and Training in Psychology. 2019. *How to Recruit and Hire Ethnic Minority Faculty.* American Psychological Association.

Committee on the Status of Minority Groups in the Economics Profession. 2019. "Report of the Committee on the Status of Minority Groups in the Economics Profession (CSMGEP)." AEA. https://www.aeaweb.org/content/file?id=9030.

Cook, Lisa D. 2019. "Mentoring Undergraduate Women Who Are Students of Color." *CSWEP News* 1: 8–10.

Darity, William A. 2010. "Notes from the Back of the Academic Bus." In *The Future of Diversity*, edited by Daniel Little and Satya P. Mohanty, 173–80. New York: Palgrave Macmillan.

Darity, William A., Jr., Darrick Hamilton, and James B. Stewart. 2015. "A Tour de Force in Understanding Intergroup Inequality: An Introduction to Stratification Economics." *The Review of Black Political Economy* 42 (1–2): 1–6.

Div.E.Q. "Videos on Economists and Their Research." 2019. Video. http://diversifyingecon.org/index.php/Videos_on_economists_and_their_research (accessed January 29, 2020).

Educational Testing Service (ETS). 2018. "A Snapshot of the Individuals Who Took the GRE General Test: July 2012–June 2017." ETS. https://www.ets.org/s/gre/pdf/snapshot_test_taker_data_2017.pdf (accessed June 25, 2020).

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos. 2014. "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom." *American Economic Review* 104 (8): 2567–91.

Francis, Dania V., Angela C. M. de Oliveira, and Carey Dimmitt. "Do School Counselors Exhibit Bias in Recommending Students for Advanced Coursework?" *The BE Journal of Economic Analysis & Policy* 19 (4).

Langin, Katie. 2019. A Wave of Graduate Programs Drops the GRE Application Requirement. *Science Magazine*. May 2019.

Lusher, Lester, Doug Campbell, and Scott Carrell. 2018. "TAs Like Me: Racial Interactions between Graduate Teaching Assistants and Undergraduates. *Journal of Public Economics* 159: 203–24.

Malmendier, Ulrike, Stefan Nagel, and Zhen Yan. 2017. "The Making of Hawks and Doves: Inflation Experiences on the FOMC." NBER Working Paper 23228.

Malveaux, Julianne. 1991. "Missed Opportunity: Sadie Tanner Mossell Alexander and the Economics Profession." *The American Economic Review* 81 (2): 307–10.

May, Ann Mari, Mary G. McGarvey, and Robert Whaples. 2014. "Are Disagreements among Male and Female Economists Marginal at Best?: A Survey of AEA Members and Their Views on Economics and Economic Policy." *Contemporary Economic Policy* 32 (1): 111–32.

Miles, Matthew B., A. Michael Huberman, and Johnny Saldaña. 2019. *Qualitative Data Analysis: A Methods Sourcebook*. 4th ed. Los Angeles: Sage Publications.

Mora, Marie T. 2019. "Best Practices in Mentoring Underrepresented Minority Women in Economics." *CSWEP News* 1: 1–4.

National Center for Science and Engineering Statistics. 2018. "Survey of Earned Doctorates. Special Tabulation, RTI International." National Science Foundation. https://www.nsf.gov/statistics/srvydoctorates/#tabs-2 (accessed February 12, 2019).

Posselt, Julie R. 2016. *Inside Graduate Admissions: Merit, Diversity, and Faculty Gatekeeping*. Cambridge, MA: Harvard University Press.

Sharpe, Rhonda Vonshay. 2017. "We've Built the Pipeline: What's the Problem and What's Next?" *CSWEP News* 2: 10–11.

Simms, Margaret C., and Charles Z. Wilson. 2020. "The National Economic Association at 50 Years: Looking Ahead." *The Minority Report* 12 (Winter): 9–11.

Spriggs, William. 2020. "Is Now a Teachable Moment for Economists? An Open Letter to Economists from Bill Spriggs." https://www.minneapolisfed.org/~/media/assets/people/william-spriggs/spriggs-letter_0609_b.pdf. (accessed June 10, 2020).

Stassun, Keivan G., Susan Sturm, Kelly Holley-Bockelmann, Arnold Burger, David J. Ernst, and Donna Webb. 2011. "The Fisk-Vanderbilt Master's-to-Ph.D. Bridge Program: Recognizing, Enlisting, and Cultivating Unrealized or Unrecognized Potential in Underrepresented Minority Students." *American Journal of Physics* 79 (4).

Tversky, Amos, and Daniel Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive psychology* 5 (2): 207–32.

Walton, Gregory M., and Geoffrey L. Cohen. 2011. "A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students." *Science* 331 (6023): 1447–51.

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 220–236

Facts and Myths about Misperceptions

Brendan Nyhan

On August 7, 2009, former vice presidential candidate Sarah Palin reshaped the debate over the Patient Protection and Affordable Care Act when she published a Facebook post falsely claiming that "my parents or my baby with Down Syndrome will have to stand in front of [Barack] Obama's 'death panel' so his bureaucrats can decide . . . whether they are worthy of health care." Palin's claim was quickly amplified by the media and in public town hall meetings with members of Congress. Within two weeks, 86 percent of Americans said they had heard "a lot" (41 percent) or "a little" (45 percent) about the myth, which three in ten people reported believing, including 47 percent of Republicans (Pew Research Center 2009). Notably, those Republicans who saw themselves as more knowledgeable about the Obama plan were significantly more likely to endorse the myth (Nyhan 2010), which persisted for years afterward.

The Affordable Care Act was ultimately enacted into law in 2010, but the "death panel" myth appeared to exert an important influence on the debate over end-of-life care. Most notably, a provision to have Medicare cover doctors' voluntary discussions with patients about end-of-life care—a policy that previously had bipartisan support—was stripped from the bill to avoid inflaming the issue further. The Obama administration later enacted this provision via regulation in 2015.

As the "death panel" myth illustrates, misperceptions threaten to warp mass opinion, undermine democratic debate, and distort public policy on issues ranging from climate change to vaccines. I define misperceptions as belief in claims that

■ *Brendan Nyhan is Professor of Government, Dartmouth College, Hanover, New Hampshire. His email is nyhan@dartmouth.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.220.

can be shown to be false (for example, that Osama bin Laden is still alive) or unsupported by convincing and systematic evidence (for example, that vaccines cause autism). In turn, I define the false or unsupported claims that help create these beliefs as misinformation. Unlike terms like "lie" and "disinformation," this approach does not require knowledge of a speaker's intent in making claims, which can rarely be established with certainty (it is unknown, for instance, whether Palin sincerely believed there would be a "death panel" or instead intended to deceive her audience). Moreover, focusing on both false and unsupported claims appropriately encompasses a great deal of dubious information—such as claims of hidden and unobserved conspiracies—that cannot be directly disproven and would be excluded by a strict insistence on falsity. However, judgment is still required about which beliefs qualify as misperceptions. For instance, some scientific findings are backed by highly credible evidence (like the role of humans in climate change), while others are more uncertain (like how certain changes in diet affect health).

In politics, the sources of—and belief in—dubious claims that meet this standard often divide along partisan lines. On the issue of health care, for instance, Politifact selected Palin's "death panel" claim as the "Lie of the Year" in 2009 and Barack Obama's oft-repeated claim that "if you like your health care plan, you can keep it" under the Affordable Care Act as the "Lie of the Year" in 2013 (Holan 2009, 2013). Public beliefs in such claims are frequently associated with people's candidate preferences and partisanship. One December 2016 poll found that 62 percent of Trump supporters endorsed the baseless claim that millions of illegal votes were cast in the 2016 election, compared to 25 percent of supporters of Hillary Clinton (Frankovic 2016). Conversely, 50 percent of Clinton voters endorsed the false claim that Russia tampered with vote tallies to help Trump, compared to only 9 percent of Trump voters. But not all political misperceptions have a clear partisan valence: for example, 17 percent of Clinton supporters and 15 percent of Trump supporters in the same poll said the US government helped plan the terrorist attacks of September 11, 2001.

Misperceptions like these often linger for years despite extensive efforts to correct the record. The same December 2016 YouGov poll found, for instance, that 36 percent of Americans think President Obama was born in Kenya, 53 percent believe there were weapons of mass destruction in Iraq that were never found, and 31 percent believe vaccines have been shown to cause autism—all claims that have been repeatedly and systematically debunked (Frankovic 2016).

In this article, I first consider the evidence that misperceptions represent genuine beliefs and are not just artifacts of question wording, partisan cheerleading, or "trolling." I then examine the psychological factors that increase vulnerability to misperceptions, especially consistency with political predispositions or group identity. Next, I turn to the sources of the false and unsupported claims that help to create and disseminate misperceptions, especially political elites and social media platforms. I then consider how misperceptions might be reduced, comparing demand-side approaches in education and journalism with supply-side interventions that try to dissuade elites from promoting misinformation or seek to limit its spread.

In each case, I argue we must carefully assess the merits of these policies rather than rushing into ill-considered responses based on media hype. Misperceptions present a serious problem, but claims that we live in a "post-truth" society with widespread consumption of "fake news" are not empirically supported and should not be used to support interventions that threaten democratic values.

Measuring Misperception Beliefs and Effects

Evidence about the prevalence of misperceptions and the characteristics of the people who hold them are typically measured using survey questions. In the past, these data were collected via phone or face-to-face interviews, but they are increasingly gathered online today. Such studies typically ask people whether they believe in or agree with various factual claims or ask them to select which statement best represents their beliefs about a disputed factual question. These findings thus have all the standard limitations of survey data such as potential sensitivity to question wording and sampling error. However, several concerns are particularly acute in the study of misconceptions.

First, available survey data is skewed toward items measuring belief in politically controversial or polarizing misperceptions. Though these misperceptions are often the most widely covered in the media, they are not necessarily representative of the set of false beliefs Americans hold, which are frequently bipartisan. For example, Graham (2020) finds that both Republicans and Democrats were (mistakenly) confident that crime and the federal budget deficit increased during the Obama presidency. Second, the survey format, which requires respondents to construct responses in a top-of-head fashion, is vulnerable to respondents reporting beliefs they did not previously hold with certainty or would not otherwise express. Graham finds that respondents are more likely to identify incorrect responses as low-certainty guesses, suggesting people often recognize what they don't know.

In addition, factual questions about politically controversial topics or figures can be vulnerable to "partisan cheerleading," which refers to providing inauthentic responses that are politically congenial. To test this conjecture, Bullock et al. (2015) and Prior, Sood, and Khanna (2015) conduct experiments among convenience samples of US respondents estimating the effect of financial incentives for accuracy on partisan polarization in factual questions about politics. Both find that the presence of incentives (payments for correct answers from $.10 to $2 or improved chances in a lottery with a payout of $200) reduced the partisan divide in expressed factual beliefs by more than 50 percent. However, these results do not necessarily indicate that people secretly hold more accurate and/or less polarized beliefs with certainty that they otherwise refuse to report. Respondents may instead use different, less error-prone guessing strategies in response to financial incentives or devote greater cognitive effort to the task than they would in real-world settings (where accuracy incentives are weak).

These problems are most severe for when people provide insincere responses (sometimes for partisan reasons) or providing insincere responses for fun or amusement ("trolling"). For example, Schaffner and Luks (2018) showed respondents pictures of the inauguration crowds from the inauguration of President Obama in 2009 and President Trump in 2017. When the pictures were unlabeled, there was broad agreement that the Obama crowd was larger, but when the pictures were labelled, many Trump supporters looked at the pictures and indicated that Trump's crowd was larger, an obviously false claim that the authors refer to as "expressive responding." Lopez and Hillygus (2018) consider the related problem of "survey trolls," which they show can inflate reported beliefs in misperceptions. This problem seems most severe for unfamiliar and outlandish rumors (for example, that Senator Ted Cruz is the Zodiac killer), where they find half or more of those indicating belief in the claim repeatedly offer unlikely responses to other questions or admit to trolling.

However, evidence suggests that surveys can provide meaningful measures of belief in prominent misperceptions. First, reported partisan differences in salient, controversial factual beliefs persist even when incentives are provided. When Peterson and Iyengar (2019) offered incentives of $.50 per question for correct answers to questions that feature ongoing factual controversy, they found that substantial partisan gaps persist—approximately two-thirds of the initial unincentivized beliefs. Similarly, Berinsky (2018) finds that providing non-financial incentives designed to reduce expressive responding (including making a survey five minutes shorter if respondents did not indicate believing in a false claim) had null effects on reported beliefs in the false claims that Barack Obama is Muslim and that the Bush administration assisted or allowed the terrorist attacks of September 11, 2001. Most recently, Allcott et al. (2020) find that financial incentives (entries in lotteries with payments of up to $100 depending on response accuracy) decrease the partisan divide in expected approval of President Trump's handling of the coronavirus outbreak at the end of April 2020 but had no measurable effect on the divide in the expected number of cases at that time, suggesting that partisans were being sincere when Republicans rated the pandemic as less severe than Democrats.

We also observe important differences in high-stakes behaviors by partisanship that are consistent with sincere (and often unsupported) differences in factual belief. First, Krupenkin (2020) finds that co-partisans of the president express more trust in vaccine safety and greater intention to vaccinate themselves and their children than opposition partisans. These patterns, which were observed during both the administrations of George W. Bush and Barack Obama, are associated with changes in real-world behavior (though medical privacy laws and ecological inference concerns limit what can be demonstrated directly). After President Obama took power, vaccination exemption rates increased differentially in Republican school districts in California compared to Democratic ones. Krupenkin, Rothschild, and Hill (2019) similarly find searches for cars and houses differentially decreased among Democrats after the 2016 election and that registrations of new cars increased less in Democratic-leaning zip codes than Republican ones. Finally,

Allcott et al. (2020) find that the individual-level partisan differences in perceived COVID-19 severity they observe in their incentivized survey data are mirrored in differences in cellphone-based measures of mobility during the pandemic between Democratic and Republican counties. These examples suggest that partisan differences in factual beliefs can affect real-world decisions and are not just cheap talk.

Less is known about the effects on misperceptions on political attitudes and policy outcomes. Misperceptions are often associated with individual-level policy and candidate preferences (for example, opponents of the Patient Protection and Affordable Care Act are more likely to believe in "death panels"), but we lack a systematic understanding for when factual beliefs are the basis for a preference versus a rationalization for a preference that a respondent would hold anyway.

To disentangle this relationship, some researchers have randomized the provision of factual information, but results from this literature are mixed. Some studies find no effect of factual corrections on related policy or candidate preferences. For example, Nyhan et al. (2019) carried out parallel experiments via Amazon Mechanical Turk and the survey research firm Morning Consult during the 2016 campaign in which respondents were randomized to view different versions of a journalistic fact-check of candidate Trump. Exposure to this information reduced misperceptions about the factual issue (in this case, changes in the prevalence of crime) but had no measurable effect on candidate support. Similarly, Hopkins et al. (2019) find across seven studies that providing information about the actual number of immigrants (which people often exaggerate) has little effect on attitudes toward immigration policy. However, other research indicates that views or preferences can change after respondents receive accurate information. For example, learning who actually pays the estate tax in a survey experiment led to increased support for the tax, especially among conservatives and Republicans with lower incomes (Sides 2016). In addition, a nationally representative survey experiment found that providing specific facts about issues like crime rates or the share of federal spending going to foreign aid affected people's policy opinions (Gilens 2001). These effects were greatest for people who were already highly knowledgeable about politics.

These results suggest that factual beliefs are not always the basis for people's policy opinions and candidate preferences. Future research should seek to develop and test cross-domain theories about the conditions under which accurate information will change people's views—for instance, is attitude change in response to new factual information more likely when partisan cues or predispositions are weak or when respondents are "cross-pressured" by competing motives?

Determining the effect of misinformation and misperceptions on media coverage and policy outcomes is an important topic but faces even more difficult theoretical and research design challenges. For example, though debates over the Patient Protection and Affordable Care Act of 2010 and end-of-life care were surely affected by misinformation, we cannot easily estimate the difference between what took place and a counterfactual version of the debate in which the "death panel" and "if you like your health plan. . ." claims were never made. Moreover, any such differences could prove to be partial equilibrium effects. If those claims had failed

to take hold, politicians and interest groups might have promoted other misperceptions instead.

Individual-level Vulnerabilities to Misperceptions

What factors make people vulnerable to believing in misperceptions? A critical and often neglected step is simple exposure. People are more likely to endorse claims to which they have been exposed—at least absent effortful resistance (Gilbert, Tafarodi, and Malone 1993). Moreover, such exposure can lead people to be more likely to endorse a claim to which they have previously been exposed even if the claim is implausible or if they possess the relevant knowledge to know that the claim is inaccurate. For example, Fazio et al. (2015) find a greater proportion of "true" ratings among undergraduates when evaluating claims like "The Atlantic Ocean is the largest ocean on Earth" if they had been randomly exposed to it before. This "illusory truth" effect seems to be the result of people using the feeling of fluency they experience when processing a familiar claim as a heuristic for truth.

These exposure effects are most likely to cumulate for people who pay more attention to potentially misleading news and information. As consumer choice has grown, differences in news consumption have widened (Prior 2005), including consumption of news from outlets that promote low-quality information. During the final weeks of the 2016 campaign, for instance, more than six in ten visits to websites that have been identified as untrustworthy by journalists and human coders came from approximately 20 percent of the US population (Guess, Nyhan, and Reifler 2020). Similarly, Pew found that the top one-third of cable news viewers average 72 minutes per day compared to three minutes and less than one minute, respectively, for the bottom two terciles (Jurkowitz and Mitchell 2013). Correspondingly, consumers of ideological and partisan news on television and online are more likely to hold misperceptions (Kull, Ramsay, and Lewis 2003; Garrett, Weeks, and Neo 2016), though establishing the direction of causality is not possible using cross-sectional observational data.

Beyond mere exposure effects, misperceptions are more likely to form and spread when people fail to apply adequate cognitive scrutiny or attention to dubious claims they encounter. One risk factor is a tendency to rely on intuitive rather than analytical thinking. Pennycook and Rand (2019) seek to evaluate this claim using performance on the Cognitive Reflection Test (CRT), a three-item battery of mathematical questions in which respondents must resist selecting an intuitive but incorrect answer and instead identify the correct answer through analytical reasoning. They found that CRT performance was correlated with the ability to distinguish between false and real news among 3,400 respondents recruited on Amazon Mechanical Turk. Similarly, reminders of accuracy (by being asked a question about whether a headline was accurate) reduced both intentions to share false news headlines that respondents could identify as false when asked and real-world sharing of information from untrustworthy websites on Twitter

(Pennycook et al. 2020). These results suggest that accuracy considerations may be given less attention by default. Finally, people may be particularly vulnerable to misinformation from trusted sources, given the way many use source identity as a heuristic for accuracy. In a study conducted using Amazon Mechanical Turk, Swire et al. (2017) find, for example, that attributing claims to Trump increased belief in them among his Republican supporters and decreased belief in them among Democrats.

People do not necessarily accept every piece of information they encounter, however. Instead, many seem especially susceptible to misperceptions that are consistent with their beliefs, attitudes, or group identity. Their psychological motivation to believe one side of a factual question seems to overwhelm their motivation to hold an accurate belief (Kunda 1990). As a result of this predisposition, which is known as "directionally motivated reasoning," we may be more skeptical of information that contradicts our preferences and more accepting of confirmatory information. Ditto and Lopez (1992) find, for example, that people who receive unwelcome medical news are more likely to question the result.

These tendencies can be especially powerful in contexts like politics where people often have strong directional preferences between parties or candidates, weak accuracy motives, and lack evidence that would resolve factual disputes.[1] Taber and Lodge (2006) find that participants were more likely to counterargue when faced with contradictory arguments about affirmative action and gun control and were more likely to accept uncritically those that reinforced their views. Such tendencies can also influence beliefs about outgroups. People are prone to hold negative false beliefs about individuals who differ from them—for example, on racial, ethnic, or religious grounds. For example, Jardina and Traugott (2019) find that belief in the "birther" myth that Barack Obama was not born in the United States was strongly associated with a survey scale measuring feelings of racial resentment among white respondents in the 2012 American National Election Study.

A particular analytical challenge is distinguishing between directionally motivated reasoning and differences in information evaluation resulting from differing priors, which are often observationally equivalent despite occurring via different processes (Druckman and McGrath 2019; Tappin, Pennycook, and Rand 2020). Isolating directionally motivated reasoning requires experimental designs that hold information fixed and manipulate processing goals. For instance, Kahan et al. (2017) presented respondents with a 2×2 table that was alternately labeled as presenting outcomes from skin cream tests (and its effect on rashes) or a ban on concealed carry for handguns (and its effect on crime). The table is designed to suggest an intuitive but false answer based on the raw totals; instead, respondents have to compute the relevant conditional probabilities to assess effectiveness.

[1] Current research seeks to propose and test models of directionally motivated reasoning showing how people deviate from the Bayesian ideal when updating their beliefs (for example, Fryer Jr., Harms, and Jackson 2018; Thaler 2020). See the recent JEP "Symposium on Motivated Beliefs" for further discussion (Epley and Gilovich 2016; Bénabou and Tirole 2016; Golman et al. 2016).

Respondents were far more polarized by partisanship and ideology over the accuracy of the test when the table concerned gun control, indicating that directional motivations influenced how the data was being processed.

Vulnerability to misinformation may also vary depending on people's knowledge and sophistication. Theoretically, being better informed might seem to protect people against holding inaccurate beliefs. However, people who are more knowledgeable are also better able to identify congenial claims and reject those that are uncongenial (Zaller 1992). In the Kahan et al. (2017) study described above, for instance, polarization in interpretation of the data depending on whether it was labeled as concerning skin cream or gun control was greatest among the most numerate respondents, who still tended to accept the intuitive answer when it was congenial but were able to figure out the correct answer when the intuitive answer was uncongenial. Similarly, Republicans who were more politically knowledgeable were *more* likely to endorse a conspiracy theory about Barack Obama manipulating unemployment statistics than less knowledgeable co-partisans (Nyhan 2012).

The Supply of Misinformation

Widespread public misperceptions often originate in dubious allegations made by prominent political figures and groups or by false rumors circulating via online or offline networks. These supply-side factors can play a critical role in the availability and salience of misinformation as well as the extent to which specific claims come to be widely believed.

Political misinformation often originates at the elite level from sources such as politicians, pundits, and ideological or partisan groups and media outlets. Though exceptions exist (for example, conspiracy theories about 9/11), elites have played a key role in creating or popularizing many of the most salient misperceptions of recent years, including the "death panel" myth and false claims that Barack Obama is a Muslim or not born in this country.

Climate change denial provides a valuable illustration of how information flows from elites can lead to widespread misperceptions. Conservatives were actually more likely than liberals to believe scientists about climate change in the 1990s before it became a partisan issue (Tesler 2018). As messages from conservative elites and Republican officials denying climate change became more widespread and salient, however, belief polarization on the issue increased (McCright and Dunlap 2011). The issue is not general ignorance: Democrats and Republicans have similar levels of knowledge about science (Kennedy and Hefferon 2019). Instead, the relationship between general scientific knowledge and belief in anthropogenic climate change now differs sharply by party. Conservative Republicans who know more about science know more about what climate scientists believe, but they simply do not endorse those claims (Kahan 2015). The most plausible mechanism for this finding is elite information flows: indeed, Tesler (2018) finds that climate denial is greatest among the conservatives with high

political interest and education who are most likely to have received the messages in question.

The incentives for political figures and groups to make such claims are clear in an era in which ideological polarization between the parties in Congress has reached historic levels (Poole and Rosenthal 2011), and partisans in the mass public express increasingly hostile feelings toward the opposition (Iyengar and Westwood 2014). Changes in media and communication have also reduced the costs of information distribution and allowed these polarized elites to communicate in a less filtered and more targeted manner with like-minded audiences (via social media, cable news, and other channels).

Economic incentives also clearly play an important role in encouraging the production of false and misleading information. Michael Moore's highly successful films "Bowling for Columbine" (2002) and "Fahrenheit 9/11" (2004), for instance, used inaccuracies and misleading innuendo that appealed to liberals who opposed George W. Bush (Nyhan 2004). Similar incentives encourage hosts on talk radio and cable news to promote misleading claims and conspiracy theories that engage and enrage their audiences. More recently, untrustworthy and "hyper-partisan" websites and Facebook pages have proliferated online (Silverman et al. 2016; Allcott and Gentzkow 2017). These outlets take advantage of the profit opportunities provided by the combination of increased demand (resulting from political polarization), low production costs (content creation without original reporting is inexpensive), and low barriers to entry and distribution on social media (which puts outlets on a more level playing field online compared to offline).

The means by which people acquire and consume information also play an important role in misperceptions. Notably, false beliefs are often attributed to the public being trapped in "echo chambers" or "filter bubbles" of politically congenial news and information online. However, the extent to which technology has created homogenous flows of information has often been overstated. Behavioral data reveal that most Americans do not have heavily slanted political information diets. For example, Guess (2018) looks at a nationally representative sample of online media use in 2015 and 2016 and finds most people pay relatively little attention to political news and/or have relatively balanced information diets, while those who frequently seek out like-minded partisan or ideological websites are a minority. Similarly, Gentzkow and Shapiro (2011) find that segregation by ideology in web-browsing data from 2004 to 2008 was modest and typically far less than people's offline networks.

However, technology may aid in the propagation of false information even if it does not create ideological or partisan segregation to the extent that critics fear. These fears have found support in studies of social media. In 2008, for instance, rumors that Barack Obama was a Muslim circulated widely online, driving up beliefs in the myth by 4–8 percentage points nationwide (Kim and Kim 2019). Correspondingly, data from 126,000 rumor cascades on Twitter from 2006 to 2017 shows that claims that were fact-checked and found to be false spread further and faster on Twitter than claims that were found to be true—a result that appears to

be attributable to the novelty of false information (Vosoughi, Roy, and Aral 2018). Low-quality websites that frequently publish false or unsupported information have sought to exploit these vulnerabilities. During the 2016 general election campaign, for instance, these sites were especially successful at using Facebook to promote their work. We observe this finding in behavioral data; Facebook was disproportionately likely to appear among the websites that Americans visited immediately prior to visiting an untrustworthy website (Guess, Nyhan, and Reifler 2020). However, these exposures again tend to be heavily concentrated. Grinberg et al. (2019) find that approximately 80 percent of false news exposures on Twitter before the 2016 election came among 1 percent of users; outside of this outlier group, they estimate that fake news sources made up only about 1 percent of the political URLs people saw on Twitter.

Reducing Misperceptions

Many observers believe that journalists and civic groups should do more to counter misperceptions in order to minimize their potentially harmful effects. It is clear that people have weak incentives to hold accurate beliefs and strong directional motivations to endorse beliefs that are consistent with a group identity such as partisanship. Conversely, political elites have strong incentives to promote misinformation and increasingly effective means of transmitting those claims to their followers. The interventions described below seek to address both of these problems.

Are such interventions necessary? One response is to argue that factual evidence ultimately wins out. Porter and Wood (2019) report the results of numerous experiments showing that people generally update their beliefs at least in part when exposed to factual information. At the macro level, Stimson and Wager (2020) point to long-term trends on high-profile issues such as the state of the economy, the link between smoking and cancer, and belief in natural selection to argue that public opinion tends to converge toward the evidence. However, the updating of beliefs may be slow (the trial of John Scopes for teaching evolution in a public school happened in 1925) and/or incomplete (on climate change, as described above). Moreover, beliefs on some issues prove to be stubbornly resistant to updating—partisans have increasingly diverged in their evaluations of the economy, for instance, since George W. Bush's presidency.

How, then, should misperceptions be reduced? Proposals that seek to address the problem vary in both the timing of the intervention (before or after exposure to or dissemination of a claim) and the target of the intervention (the public, political elites, or social media platforms).[2] This typology is summarized in Table 1. I briefly review the evidence for each approach below.

[2]I am indebted to Andy Guess for this point.

Table 1
Target and Timing of Interventions to Reduce Misperceptions

	Individuals	*Political elites*	*Online platforms*
In advance	Political information/ media literacy	Reputational incentives	Reduce untrustworthy sources
Afterward	Corrections/ fact-checking	Reputational sanctions	Fact-check labels/ reduce reach

One way to prevent misperceptions, some argue, is for journalists, educators, and other nonpartisan organizations and institutions to provide people with more or better information about the issues in question in advance. However, the evidence supporting this conjecture is mixed.

First, we lack a social consensus on the institutions that would provide such information. The problem is not a lack of capacity. Though failures of course exist, the United States and other industrialized countries generally have well-functioning government agencies such as the Bureau of Labor Statistics and the Centers for Disease Control and Prevention that provide accurate information about metrics such as unemployment or health care outcomes. Similarly, though journalism is an inherently subjective enterprise, research suggests that the conclusions of fact-checking sites—the media outlets most closely aligned with this mission—are generally aligned for claims that are rated as clearly true or false, though agreement is less consistent for claims between those endpoints (Amazeen 2016; Lim 2018). Nonetheless, both government statistics and the media are widely distrusted, especially among Republicans (Frankovic 2017; Guess, Nyhan, and Reifler 2019).

In addition, credible civic and political information may fail to attract public attention, particularly from voters who might need it most. Fewer than half of the Americans who were exposed to news from untrustworthy websites even visited a fact-checking website in the weeks before the 2016 election (Guess, Nyhan, and Reifler 2020). In general, people tend to prefer other kinds of content. Iyengar, Norpoth, and Hahn (2004) tested preferences for news about the 2000 election and found that people tended to prefer coverage of the horse race and political strategy to factual information about issues. The people who prefer factual information are not typical. When Mummolo and Peterson (2017) looked at a voter guide produced for the *Sacramento Bee* newspaper in 2014, for example, they found that it was mostly used by people who are already highly interested in and knowledgeable about politics. It is thus unclear that providing more political information or improving political knowledge will reduce the prevalence of misperceptions.

An alternate approach is to build "media literacy," which seeks to help people better identify (un)trustworthy information sources on their own. Experimental studies suggest that even brief exposure to interventions that provide guidelines and recommendations for identifying accurate information can reduce belief in false claims and help people distinguish between false and mainstream news. For

example, reading the "tips" for spotting untrustworthy news provided by Facebook and WhatsApp increased participant discernment between mainstream and false news headlines in studies conducted in the United States and India (Guess et al. forthcoming). Roozenbeek and van der Linden (2019) also found that the experience of playing a "fake news" game in which users learned misinformation tactics helped them better identify unreliable headlines and tweets afterward. However, most media literacy interventions have not yet been evaluated in randomized trials. Moreover, even if these efforts prove effective, they may be difficult and/or costly to implement and scale in a manner that creates durable effects, especially outside the education system.

Fact-checks instead seek to counter misinformation by evaluating the accuracy of claims directly, including after they are made. For example, fact-checkers might seek to debunk false or misleading claims to which people have been exposed after a political debate or presidential address. An early study in this literature found evidence of a "backfire effect" in which people who were exposed to counter-attitudinal corrective information then expressed more belief in a misconception (Nyhan and Reifler 2010). However, this finding appears to have been anomalous (Nyhan forthcoming). Meta-analyses of the related literatures on corrective information and fact-checking find that they do generally increase the accuracy of people's beliefs and reduce belief in misperceptions, though these interventions do not fully offset the effect of exposure to misinformation and their effects may be reduced in conflictual political settings (Chan et al. 2017; Walter and Murphy 2018; Walter et al. 2019).

Moreover, post-exposure fact-checks share some of the same problems as efforts to provide accurate information in advance of exposure. As noted above, articles on fact-checking websites are poorly targeted to people who are exposed to the misinformation those articles seek to debunk (Guess, Nyhan, and Reifler 2020). In addition, the effects of fact-check exposure tend to decay over time. These decay effects may be larger when high-profile politicians or issues are involved. In a study of respondents from Amazon Mechanical Turk, for instance, Swire et al. (2017) find that increases in belief accuracy after affirmations of true statements or fact-checks of misinformation declined after a week and that these effects were larger when the claim in question was attributed to Donald Trump. These findings may help to explain why the encouraging results seen in many one-shot fact-check experiments do not translate into sustained reductions in belief in high-profile misperceptions even when corrective information is widely disseminated.

In general, interventions targeting the public face difficult issues in reaching the individuals who hold misperceptions, creating durable changes in beliefs, and scaling in a cost-effective manner across the population. It is therefore important to also consider alternate approaches that seek to limit misperceptions by reducing the supply of misinformation and its spread.

One approach is to change the incentives or practices of political elites and publishers. In one field experiment testing the effects of these incentives, a random subset of state legislators from nine states were sent messages before the 2012

election about the political costs of having false claims identified by fact-checkers. Those who were sent the messages were less likely to have the accuracy of their statements questioned publicly, suggesting that the reminder discouraged false claims (Nyhan and Reifler 2015). Facebook has also announced that it would reduce the reach of groups that repeatedly post false claims and content from publishers who try to game Facebook's algorithms but have limited reach online, which may not only reduce the prevalence of misinformation but discourage publishers from using such tactics (Dreyfuss and Lapowsky 2019).

In addition, online platforms can warn people about false claims and limit their reach when they have been identified by third-party fact-checkers, overcoming the scale and targeting problems that fact-checkers otherwise face. Facebook has made the most extensive efforts in this regard and has seemingly succeeded in reducing the prevalence of false content in the News Feed. Guess et al. (2018) estimate that visits to untrustworthy websites by Americans declined from 27 percent in fall 2016 to 7 percent in fall 2018. Allcott, Gentzkow, and Yu (2019) also find a differential decline in fake news stories during this period on Facebook relative to Twitter, which employs less aggressive content moderation practices, suggesting the same conclusion.

Conclusion

Many responses to the problem of misinformation unfortunately threaten to undermine or limit free speech in democratic societies. For example, critics have called on Facebook to ban ads from political candidates that are deemed false, which would introduce a centralized constraint on a core form of political speech that is absent in other media like television. Since 2016, a number of countries around the world have gone even further in using fines or even criminal penalties to try to limit misinformation. For example, Kenya enacted legislation making the publication of false information a crime, a step that the Committee to Project Journalists said will criminalize free speech (Malalo and Mohammed 2018).

Calls for such draconian interventions are commonly fueled by a moral panic over claims that "fake news" has created a supposedly "post-truth" era. These claims falsely suggest an earlier fictitious golden age in which political debate was based on facts and truth. In reality, false information, misperceptions, and conspiracy theories are general features of human society. For instance, belief that John F. Kennedy was killed in a conspiracy were already widespread by the late 1960s and 1970s (Bowman and Rugg 2013). Hofstadter (1964) goes further, showing that a "paranoid style" of conspiratorial thinking recurs in American political culture going back to the country's founding. Moreover, exposure to the sorts of untrustworthy websites that are often called "fake news" was actually quite limited for most Americans during the 2016 campaign—far less than media accounts suggest (Guess, Nyhan, and Reifler 2020). In general, no systematic evidence exists to demonstrate that the prevalence of misperceptions today (while worrisome) is worse than in the past.

Even exposure to the ill-defined term "fake news" and claims about its prevalence can be harmful. In an experimental study among respondents from Mechanical Turk, Van Duyn, and Collier (2019) find that when people are exposed to tweets containing the term "fake news," they become less able to discern real from fraudulent news stories. Similarly, Clayton et al. (2019) find that participants from Mechanical Turk who are exposed to a general warning about the prevalence of misleading information on social media then tend to rate headlines from both legitimate and untrustworthy news sources as less accurate, suggesting that the warning causes an indiscriminate form of skepticism.

Any evidence-based response to the problem of misperceptions must thus begin with an effort to counter misinformation about the problem itself. Only then can we design interventions that are proportional to the severity of the problem and consistent with the values of a democratic society.

■ *I thank the Carnegie Corporation of New York and the National Science Foundation (award 1949077) for financial support. I am grateful to Ben Lyons, Shun Yamaya, and the editors for helpful comments and to my co-authors Andy Guess, Ben Lyons, Jacob Montgomery, and Jason Reifler for the joint work that has shaped my thinking on these topics.*

References

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211–36.

Allcott, Hunt, Matthew Gentzkow, and Chuan Yu. 2019. "Trends in the Diffusion of Misinformation on Social Media." *Research & Politics* 6 (2): 1–8.

Allcott, Hunt, Levi Boxell, Jacob C. Conway, Matthew Gentzkow, Michael Thaler, and David Y. Yang. 2020. "Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic." NBER Working Paper 26946.

Amazeen, Michelle A. 2016. "Checking the fact-checkers in 2008: Predicting political ad scrutiny and assessing consistency." *Journal of Political Marketing* 15(4): 433-464.

Bénabou, Roland and Jean Tirole. 2016. "Mindful Economics: The Production, Consumption, and Value of Beliefs." *Journal of Economic Perspectives* 30(3): 141–64.

Berinsky, Adam J. 2018. "Telling the Truth about Believing the Lies? Evidence for the Limited Prevalence of Expressive Survey Responding." *Journal of Politics* 80(1): 211–24.

Bowman, Karlyn, and Andrew Rugg. 2013. "Public Opinion on Conspiracy Theories." *AEI Paper & Studies* Novermber 2013.

Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory Huber. 2015. "Partisan Bias in Factual Beliefs about Politics." *Quarterly Journal of Political Science* 10 (4): 519–78.

Chan, Man-pui Sally, Christopher R. Jones, Kathleen Hall Jamieson, and Dolores Albarracín. 2017. "Debunking: A Meta-analysis of the Psychological Efficacy of Messages Countering Misinformation." *Psychological Science* 28 (11): 1531–46.

Clayton, Katherine, Spencer Blair, Jonathan A. Busam, Samuel Forstner, John Glance, Guy Green, Anna

Kawata et al. 2019. "Real Solutions for Fake News? Measuring the Effectiveness of General Warnings and Fact-Check Tags in Reducing Belief in False Stories on Social Media." *Political Behavior* https://doi.org/10.1007/s11109-019-09533-0.

Ditto, Peter H., and David F. Lopez. 1992. "Motivated Skepticism: Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions." *Journal of Personality and Social Psychology* 63 (4): 568–84.

Dreyfuss, Emily, and Issie Lapowsky. 2019. "Facebook Is Changing News Feed (Again) to Stop Fake News." *Wired*, April 10. https://www.wired.com/story/facebook-click-gap-news-feed-changes/.

Druckman, James N., and Mary C. McGrath. 2019. "The Evidence for Motivated Reasoning in Climate Change Preference Formation." *Nature Climate Change* 9: 111–19.

Epley, Nicholas and Thomas Gilovich. 2016. "The Mechanics of Motivated Reasoning." *Journal of Economic Perspectives* 30(3): 133–40.

Fazio, Lisa K., Nadia M. Brashier, B. Keith Payne, and Elizabeth J. Marsh. 2015. "Knowledge Does Not Protect against Illusory Truth." *Journal of Experimental Psychology General* 144 (5): 993–1002.

Frankovic, Kathy. 2016. "Belief in Conspiracies Largely Depends on Political Identity." *YouGov*, December 27. https://today.yougov.com/topics/politics/articles-reports/2016/12/27/belief-conspiracies-largely-depends-political-iden.

Frankovic, Kathy. 2017. "Does the Public Believe in Government Statistics? It Depends." *YouGov*, March 23. https://today.yougov.com/topics/politics/articles-reports/2017/03/23/does-public-believe-government-statistics-depends.

Fryer Roland G., Jr., Philipp Harms, and Matthew O. Jackson. 2018. "Updating Beliefs When Evidence Is Open to Interpretation: Implications for Bias and Polarization." *Journal of the European Economic Association* 17 (5): 1470–1501.

Garrett, R. Kelly, Brian E. Weeks, and Rachel L. Neo. 2016. "Driving a Wedge between Evidence and Beliefs: How Online Ideological News Exposure Promotes Political Misperceptions." *Journal of Computer-Mediated Communication* 21(5): 331–48.

Gentzkow, Matthew, and Jesse M. Shapiro. 2011. "Ideological Segregation Online and Offline." *Quarterly Journal of Economics* 126 (4): 1799–1839.

Gilbert, Daniel T., Romin W. Tafarodi, and Patrick S. Malone. 1993. "You Can't Not Believe Everything You Read." *Journal of Personality and Social Psychology* 65 (2): 221–33.

Gilens, Martin. 2001. "Political Ignorance and Collective Policy Preferences." *American Political Science Review* 95 (2): 379–96.

Golman, Russell, George Loewenstein, Karl Ove Moene, and Luca Zarri. 2016. "The Preference for Belief Consonance." *Journal of Economic Perspectives* 30(3): 165–88.

Graham, Matthew H. 2020. "Self-awareness of Political Knowledge." *Political Behavior* 42: 305–26.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. "Fake News on Twitter during the 2016 US Presidential Election." *Science* 363 (6425): 374–78.

Guess, Andrew, Benjamin Lyons, Brendan Nyhan, and Jason Reifler. 2018. "Fake News, Facebook Ads, and Misperceptions: Assessing Information Quality in the 2018 U.S. Midterm Election Campaign." Downloaded November 5, 2019 from http://www.dartmouth.edu/~nyhan/fake-news-2018. pdf.

Guess, Andrew, Brendan Nyhan, and Jason Reifler. 2019. "National News, Local Lens? Findings from the 2019 Poynter Media Trust Survey." October 21, 2019. Downloaded May 20, 2020 from http://www.dartmouth.edu/~nyhan/media-trust-report-2019.pdf.

Guess, Andrew, Brendan Nyhan, and Jason Reifler. 2020. "Exposure to Untrustworthy Websites in the 2016 U.S. Presidential Campaign." *Nature Human Behaviour* 4: 472–80.

Guess, Andrew M. 2018. "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets." Unpublished.

Guess, Andrew M., Michael Lerner, Benjamin Lyons, Jacob M. Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. Forthcoming. "A Digital Media Literacy Intervention Increases Discernment between Mainstream and False News in the United States and India." *Proceedings of the National Academy of Sciences*.

Hofstadter, Richard. 1964. *The Paranoid Style in American Politics*. New York: Random House.

Holan, Angie Drobnic. 2009. "PolitiFact's Lie of the Year: 'Death Panels.'" *Politifact*, December 18. https://www.politifact.com/article/2009/dec/18/politifact-lie-year-death-panels/.

Holan, Angie Drobnic. 2013. "Lie of the Year: 'If You Like Your Health Care Plan, You Can Keep It.'" *Politifact*, December 12. https://www.politifact.com/article/2013/dec/12/lie-year-if-you-like-your-health-care-plan-keep-it/.

Hopkins, Daniel J., John Sides, and Jack Citrin. 2019. "The Muted Consequences of Correct Information about Immigration." *Journal of Politics* 81 (1): 315–20.

Iyengar, Shanto, Helmut Norpoth, and Kyu S. Hahn. 2004. "Consumer Demand for Election News: The Horserace Sells." *Journal of Politics* 66 (1): 157–75.

Iyengar, Shanto, and Sean J. Westwood. 2014. "Fear and Loathing across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59 (3): 690–707.

Jardina, Ashley, and Michael Traugott. 2019. "The Genesis of the Birther Rumor: Partisanship, Racial Attitudes, and Political Knowledge." *Journal of Race, Ethnicity and Politics* 4 (1): 60–80.

Jurkowitz, Mark, and Amy Mitchell. 2013. "How Americans Get TV News at Home." *Pew Research Center*, October 11. https://www.journalism.org/2013/10/11/how-americans-get-tv-news-at-home/.

Kahan, Dan M. 2015. "Climate-Science Communication and the Measurement Problem." *Political Psychology* 36 (S1): 1–43.

Kahan, Dan M., Ellen Peters, Erica Cantrell Dawson, and Paul Slovic. 2017. "Motivated Numeracy and Enlightened Self-Government." *Behavioural Public Policy* 1 (1): 54–86.

Kim, Jin Woo, and Eunji Kim. 2019. "Identifying the Effect of Political Rumor Diffusion Using Variations in Survey Timing." *Quarterly Journal of Political Science* 14 (3): 293–311.

Krupenkin, Masha. 2020. "Does Partisanship Affect Compliance with Government Recommendations?" *Political Behavior*.

Krupenkin, Masha, David Rothschild, and Shawndra Hill. 2019. "Do Partisans Make Riskier Financial Decisions When Their Party Is in Power?" Unpublished.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108 (3): 480–98.

Kull, Steven, Clay Ramsay, and Evan Lewis. 2003. "Misperceptions, the Media, and the Iraq War." *Political Science Quarterly* 118 (4): 569–98.

Lim, Chloe. 2018. "Checking How Fact-checkers Check." *Research & Politics* July–Septemeber 2018: 1–7.

Lopez, Jesse, and D. Sunshine Hillygus. 2018. "Why So Serious?: Survey Trolls and Misinformation." Unpublished.

Kennedy, Brian, and Meg Hefferon. 2019. "What Americans Know about Science." *Pew Research Center*, March 28. https://www.pewresearch.org/science/2019/03/28/what-americans-know-about-science/.

Malalo, Humphrey, and Omar Mohammed. 2018. "Kenya's President Signs Cybercrimes Law Opposed by Media Rights Groups." *Reuters*, May 16. https://www.reuters.com/article/us-kenya-lawmaking/kenyas-president-signs-cybercrimes-law-opposed-by-media-rights-groups-idUSKCN1IH1KX.

McCright, Aaron M., and Riley E. Dunlap. 2011. "The Politicization of Climate Change and Polarization in the American Public's Views of Global Warming, 2001–2010." *The Sociological Quarterly* 52 (2): 155–94.

Mummolo, Jonathan, and Erik Peterson. 2017. "How Content Preferences Limit the Reach of Voting Aids." *American Politics Research* 45 (2): 159–85.

Nyhan, Brendan. 2004. "Fahrenheit 9/11: The Temperature at Which Michael Moore's Pants Burn." *Spinsanity*, July 2. http://www.spinsanity.org/columns/20040702.html.

Nyhan, Brendan. 2010. "Why the 'Death Panel' Myth Won't Die: Misinformation in the Health Care Reform Debate." *The Forum* 8 (1).

Nyhan, Brendan. 2012. "Political Knowledge Does Not Guard Against Belief in Conspiracy Theories." *YouGov*, November 5. http://today.yougov.com/news/2012/11/05/political-knowledge-does-not-guard-against-belief-/.

Nyhan, Brendan. 2018. "Fake News and Bots May Be Worrisome, but Their Political Power Is Overblown." *New York Times*, February 13. https://www.nytimes.com/2018/02/13/upshot/fake-news-and-bots-may-be-worrisome-but-their-political-power-is-overblown.html.

Nyhan, Brendan. Forthcoming. "'Why 'Backfire Effects' Do Not Explain the Durability of Political Misperceptions." *Proceedings of the National Academy of Sciences*.

Nyhan, Brendan, Ethan Porter, Jason Reifler, and Thomas J. Wood. 2019. "Taking Fact-Checks Literally but Not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability." *Political Behavior*.

Nyhan, Brendan, and Jason Reifler. 2010. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32 (2): 303–30.

Nyhan, Brendan, and Jason Reifler. 2015. "The Effect of Fact-Checking on Elites: A Field Experiment on U.S. State Legislators." *American Journal of Political Science* 59 (3): 628–40.

Pennycook, Gordon, and David G. Rand. 2019. "Lazy, Not Biased: Susceptibility to Partisan Fake News

Is Better Explained by Lack of Reasoning Than by Motivated Reasoning." *Cognition* 188: 39–50.

Pennycook, Gordon, Ziv Epstein, Mohsen Mosleh, Antonio Arechar, Dean Eckles, and David Rand. 2020. "Understanding and Reducing the Spread of Misinformation Online." Unpublished.

Peterson, Erik, and Shanto Iyengar. 2019. "Partisan Gaps in Political Information and Information-Seeking Behavior: Motivated Reasoning or Cheerleading?" Unpublished.

Pew Research Center. 2009. "Health Care Reform Closely Followed, Much Discussed." August 20, *Pew Research Center*.

Poole, Keith T., and Howard Rosenthal. 2011. *Ideology and Congress*. New Brunswick, NJ: Transaction Publishers.

Porter, Ethan, and Thomas J. Wood. 2019. *False Alarm: The Truth About Political Mistruths in the Trump Era*. Cambridge: Cambridge University Press.

Prior, Markus. 2005. "News vs. Entertainment: How Increasing Media Choice Widens Gaps in Political Knowledge and Turnout." *American Journal of Political Science* 49(3): 577–92.

Prior, Markus, Gaurav Sood, and Kabir Khanna. 2015. "You Cannot Be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions." *Quarterly Journal of Political Science* 10 (4): 489–518.

Roozenbeek, Jon, and Sander van der Linden. 2019. "Fake News Game Confers Psychological Resistance against Online Misinformation." *Palgrave Communications* 5.

Schaffner, Brian F., and Samantha Luks. 2018. "Misinformation or Expressive Responding? What an Inauguration Crowd Can Tell Us about the Source of Political Misinformation in Surveys." *Public Opinion Quarterly* 82 (1): 135–47.

Sides, John. 2016. "Stories or Science? Facts, Frames, and Policy Attitudes." *American Politics Research* 44 (3): 387–414.

Silverman, Craig, Lauren Strapagiel, Hamza Shaban, Ellie Hall, and Jeremy Singer-Vine. 2016. "Hyperpartisan Facebook Pages Are Publishing False And Misleading Information At An Alarming Rate." *Buzzfeed*, October 20. https://www.buzzfeednews.com/article/craigsilverman/partisan-fb-pages-analysis.

Stimson, James A., and Emily M. Wager. 2020. *Converging on Truth: A Dynamic Perspective on Factual Debates in American Public Opinion*. Cambridge: Cambridge University Press.

Swire, Briony, Adam J. Berinsky, Stephan Lewandowsky, and Ullrich K.H. Ecker. 2017. "Processing Political Misinformation: Comprehending the Trump Phenomenon." *Royal Society Open Science* 4 (3).

Taber, Charles S., and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50 (3): 755–69.

Tappin, Ben M., Gordon Pennycook, and David G. Rand. 2020. "Thinking Clearly about Causal Inferences of Politically Motivated Reasoning: Why Paradigmatic Study Designs Often Prevent Causal Inference." *Current Opinion in Behavioral Sciences* 34: 81–87.

Tesler, Michael. 2018. "Elite Domination of Public Doubts about Climate Change (Not Evolution)." *Political Communication* 35 (2): 306–26.

Thaler, Michael. 2020. "The 'Fake News' Effect: An Experiment on Motivated Reasoning and Trust in News." Unpublished.

Van Duyn, Emily, and Jessica Collier. 2019. "Priming and Fake News: The Effects of Elite Discourse on Evaluations of News Media." *Mass Communication and Society* 22 (1): 29–48.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380): 1146–51.

Walter, Nathan, Jonathan Cohen, R. Lance Holbert, and Yasmin Morag. 2019. "Fact-Checking: A Meta-Analysis of What Works and for Whom." *Political Communication* 37 (3).

Walter, Nathan, and Sheila T. Murphy. 2018. "How to Unring the Bell: A Meta-analytic Approach to Correction of Misinformation." *Communication Monographs* 85 (3): 423–41.

Wood, Thomas, and Ethan Porter. 2019. "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence." *Political Behavior* 41: 135–63.

Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 237–261

Venture Capital's Role in Financing Innovation: What We Know and How Much We Still Need to Learn

Josh Lerner and Ramana Nanda

Venture capital is associated with some of the most high-growth and influential firms in the world. For example, among publicly traded firms worldwide, seven of the top eight firms by market capitalization in May 2020 had been backed by venture capital prior to their initial public offerings: Alphabet, Apple, Amazon, Facebook, and Microsoft in the United States, and Alibaba and Tencent in China. More generally, although firms backed by venture capital comprise less than 0.5 percent of firms that are born each year in the United States, they represent nearly half of entrepreneurial companies that graduate to the public marketplace.

Academics and practitioners have effectively articulated the strengths of the venture model. These include its strong emphasis on governance by venture capital investors through staged financing, contractual provisions, and active involvement with their portfolio companies. Indeed, Kenneth Arrow (1995) once opined that "venture capital has done much more, I think, to improve efficiency than anything."

In many respects, the venture capital industry appears to be a bright spot in an increasingly troubled global innovation landscape (Bloom et al. 2020). Over the last decade, the amount of capital deployed worldwide by venture capital investors and the number of startups receiving funding have grown substantially. Entirely new financial intermediaries such as accelerators, crowdfunding platforms, and "super angels" have emerged at the early stage of new venture finance. Meanwhile, mutual

■ *Josh Lerner is Jacob H. Schiff Professor and Ramana Nanda is Sarofim-Rock Professor, both at Harvard Business School, Boston, Massachusetts. During the 2019–2021 academic years, Nanda is a Visiting Professor at Imperial College, London, United Kingdom. Both authors are also Research Associates at the National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are jlerner@hbs.edu and RNanda@hbs.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.237.

funds, hedge funds, corporations, and sovereign wealth funds have deployed large sums of capital into more mature, but still private, venture capital-backed firms.

In this paper, we acknowledge the power of the venture capital in fomenting innovation. At the same time, despite the optimism articulated by Arrow and by many other academics and practitioners, we argue that venture capital financing also has real limitations in its ability to advance substantial technological change. While our ability to assess the social welfare impact of venture capital remains nascent, we hope that this discussion will stimulate discussion and research about these questions.

Three issues are particularly concerning to us: 1) the very narrow band of technological innovations that fit the requirements of institutional venture capital investors; 2) the relatively small number of venture capital investors who hold and shape the direction of a substantial fraction of capital that is deployed into financing radical technological change; and 3) the relaxation in recent years of the intense emphasis on corporate governance by venture capital firms. We believe these phenomena, rather than being short-run anomalies associated with the ebullient equities market from the decade or so up through early 2020, may have ongoing and detrimental effects on the rate and direction of innovation in the broader economy.

We begin this paper by tracing the growth of the venture capital industry over the past 40 years, noting how technological and institutional changes have narrowed the focus and concentrated the capital invested by venture capital firms as well as potentially contributing to a decline in governance. We then turn to some potential adaptations to the venture capital industry model that might enable a broader base of ideas and technologies to receive risk capital. In particular, we propose some possibilities for altering what seems to be a standard and inflexible contract between venture capital funds and their investors as well as potential approaches to manage venture investments in certain industries more effectively.

Our focus here will primarily be on the US venture capital industry. But we would be remiss if we did not note that the growth rate of global venture capital has exceeded that of the US economy. The National Venture Capital Association (2020) estimates, for instance, that the US share of world venture capital financing has fallen from about 80 percent in 2006–2007 to under 50 percent in 2016–2019.

A Brief Look at the Development of Institutional Venture Capital

Origins

Entrepreneurs sought funds to pursue their risky ideas for centuries before the modern venture capital industry emerged. Indeed, some of the key elements of today's venture capital industry, such as the use of risk-sharing partnerships, can be traced as far back as Genoese merchants in the 15th century and American whaling voyages in the 19th century (Astuti 1933; de Roover 1963; Lopez and Raymond 1955; Nicholas 2019).

Most business historians, however, trace the origins of the institutional venture capital industry to 1946, when Harvard Business School professor Georges Doriot

formed the American Research & Development Corporation with local Boston civic leaders to invest in young ventures developed during World War II.[1] Doriot articulated and practiced many of the key principles of venture investment that continue to this day. These guideposts include: the intensive scrutiny (and frequent rejection) of business plans prior to financing, the provision of oversight as well as capital, the staged financing of investments, and the ultimate return of capital and profits to the outside investors that provided the original funding.[2]

Several of the most prominent venture capital firms of today—such as Sequoia Capital, Kleiner Perkins, and New Enterprise Associates—were formed in the early to mid-1970s to invest in what would become the burgeoning semiconductor and computer industries. However, the industry did not take off until the early 1980s, when pension funds began to allocate some of their capital towards this relatively new asset class. Much of this change can be traced back to a clarification of an obscure rule in the Employment Retirement Income Security Act (ERISA). The rule had originally stated that private pension managers had to invest their funds' resources with the care of a "prudent man," which was interpreted as requiring only very low-risk investments. In early 1979, the Department of Labor ruled that pension fund managers could take portfolio diversification into account in determining prudence, which implied that the government would not view allocation of a small fraction of a corporate pension fund portfolio to illiquid funds like venture capital as imprudent, even if a number of companies in the venture capitalist's portfolio failed. While the allocations of corporate pension funds to venture capital in the 1980s were initially very modest, even a small allocation of such a large pool led to very rapid growth of the venture capital sector.

A decade later, US public pension funds also started investing in venture capital firms and were soon followed by pension and sovereign funds from around the globe. Initially, neither private nor public pension funds invested in a dramatically different manner than their predecessors. But their impact was important because of their sheer size, which dwarfed that of the early venture capital investors such as university endowments and insurance companies. In the subsequent 40 years, venture capital has come to be established as the dominant source of financing for high-potential startups commercializing risky new ideas and technologies.

Venture Capital's Impact

Table 1 highlights how venture capital is involved in financing startups that ultimately have become some of the largest and most successful firms in the economy. We looked at the 4,109 initial public offerings over the 1995–2018 period of

[1] Our reference to "institutional" venture capital refers to the majority of the venture capital industry that raises money from and invests on behalf of "limited partners"—entities such as university endowments or pension funds that allocate some of their capital to the venture capital asset class. However, corporations, family offices, and pension funds also make direct investments into high-risk ventures.

[2] The reader desiring a more detailed perspective can study the several volumes on the industry's evolution, including Ante (2008) and Nicholas (2019). In an earlier *JEP* article, Gompers and Lerner (2001b) review the industry's first half century.

Table 1

Comparison of Publicly Traded Firms in the United States, Based on Whether Backed by Institutional Venture Capital Investors

	VC-Backed IPOs	*All IPOs*	*VC-Backed as a % of all*
Total number of non-financial IPOs between 1995 and 2019	1,930	4,109	47.0%
Number of firms still public at 12/31/2019	582	1,044	55.7%
Share of IPOs that were still public at 12/31/2019	30%	25%	
Key statistics as of December 31, 2019 for firms still public (all figures millions USD, except number of employees)			
Total enterprise value	4,844,717	7,129,838	67.9%
Total market capitalization	4,922,394	6,462,409	76.2%
Global employees	2,279,715	5,336,394	42.7%
Total revenue	1,157,679	2,171,239	53.3%
Net income	53,082	98,554	53.9%
R&D expenditure	148,388	167,442	88.6%

Source: IPO data from SDC Platinum (accessed 01/08/2020); company-level stastics from Standard and Poor's Capital IQ (accessed 04/24/2020)

Note: This table reports statistics for the sample of publicly traded firms that had an initial public offering (IPO) between 1995 and 2018 and were still public on December 31, 2019, further conditioning on those that were founded after 1980 and were not financial firms. It compares statistics for firms that were backed by venture capital firms prior to their IPO with those that were not. IPO data are drawn from Refinitiv's SDC Platinum database, with data for key statistics drawn from S&P's Capital IQ database. All attributes are measured as of December 31, 2019.

nonfinancial firms that were founded in 1980 or later. Table 1 (inspired by Gornall and Strebulaev 2015) shows that 47 percent (or 1,930) of these firms were backed by venture capital investors prior to their initial public offering. Of those 4,109 IPOs, 1,044 were still publicly traded at the end of 2019. The table compares the 1,044 firms—at the end of 2019—based on whether they were originally venture capital-backed (582 firms) versus not (462 firms). That is, 56 percent of the firms that had initial public offerings from 1995 to 2018 and were still alive at the end of 2019 were backed by venture capital. Considering that under 0.5 percent of firms in the economy receive venture capital financing (Puri and Zarutskie 2012), Table 1 highlights the disproportionate role firms backed by venture capital play in the US economy.

An important question relates to whether these differences arise purely due to venture capital firms selecting high-growth opportunities or whether these investors also play a causal role in improving the growth and performance of new companies. Discerning causality in this setting is tough, and much of the research has consequentially been more descriptive in nature. Chemmanur, Krishnan, and Nandy (2011) and Puri and Zarutskie (2012) examine the universe of firms using the Longitudinal Research Database of the US Census Bureau. They argue that the evidence is consistent with the proposition that venture capital increases firm sales and lowers the likelihood of firm failure.

Other papers have attempted to exploit discontinuities to identify the relationship between venture capital and innovation. Kortum and Lerner (2000) use the 1979 "prudent man" change in pension fund rules that increased venture capital funding as a natural experiment, along with several other approaches, to look for causality. They find that a rise in venture capital causes higher rates of patenting. Bernstein, Giroud, and Townsend (2016) examine the opening up of new airline routes that make it easier for a venture capital firm to visit one of its existing portfolio companies. They find that when it becomes easier for the venture capital firm to monitor, the portfolio firm performs better.

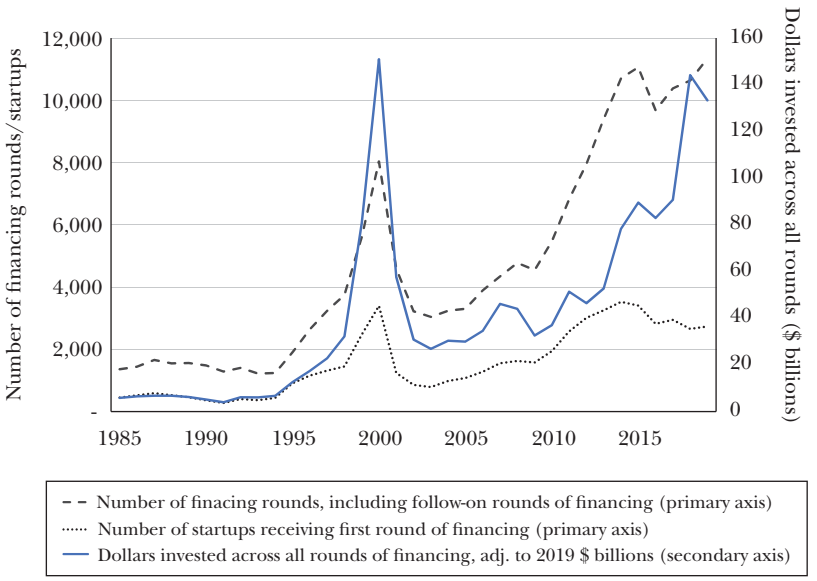
Other research has fleshed out the mechanisms that venture capital investors use. These tools include staged financing (Gompers 1995; Neher 1999), securities that have state-contingent cash flow and control rights (Hellmann 1998; Cornelli and Yosha 2003; Kaplan and Strömberg 2003, 2004), and the active role of venture capital investors on boards of portfolio companies (Hellmann and Puri 2000, 2002; Lerner 1995). The authors argue these approaches have an important impact on the success of portfolio companies. For a more thorough review of the extensive literature on venture capital, a useful starting point is Da Rin, Hellmann, and Puri (2013).

Looking back at Table 1, one can see that firms that were backed by venture capital prior to their initial public offering represent a similar share of revenues and profits as their proportion of surviving public firms (about 55 percent). On average, they are less labor intensive, more valuable in terms of market capitalization, and represent 89 percent of the recorded research and development expenditures by these firms in 2019.

The disproportionate share of recorded research and development expenditures by firms that were formerly backed by venture capital stems from two elements. First, 91 percent of public firms originally backed by venture capital recorded expenses related to research and development in 2019, compared to a much smaller 72 percent for firms not originally backed by venture capital. Second, among those that did report research and development expenses, the intensity of research and development, measured as a share of firm revenue, was higher.

The research and development intensity of publicly traded firms that were backed by venture capital relates to the role of venture capital in financing repeated waves of technological innovation: the semiconductor revolution and diffusion of mainframe computing in the 1960s; the advent of personal computing in the early 1980s; the biotechnology revolution of the 1980s; and the introduction of the Internet and e-commerce in the 1990s (which as seen from Figure 1, led to a sharp expansion in venture capital and venture capital-backed startups between 1995 and 2000). After a significant decline in 2000 and the subsequent "dot com bust," the role of venture capital firms in financing technological revolutions continued in the 2000s, as exemplified by the widespread diffusion of "smart" mobile communications technologies and new businesses enabled by the rise of cloud computing. Consistent with this pattern, Howell et al. (2020) use US patent data over the 1976–2017 period to document that venture capital-backed firms were between two

Figure 1
Evolution of the US Venture Capital Industry from 1985–2019



Source: Data are drawn from the National Venture Capital Association's Yearbooks and related publications.
Note: This figure reports the number of unique startups headquartered in the United States receiving an initial round of financing from institutional venture capital investors (left-hand axis), the total number of financing rounds associated with these startups, including follow-on rounds of financing (left-hand axis) and the total dollars invested across these rounds of financing in constant 2019 billions of dollars (right-hand axis) for each year from 1985–2019.

and four times as likely to have filed patents that were in the top percentiles of influence (as measured by citations, originality, generality, and closeness to science).

Venture Capital in the 2010s

Venture capital has boomed over the past decade, driven by new investment opportunities and greater availability in the supply of capital for this asset class.

On the demand side, there has been a plethora of attractive investment opportunities. Many venture-backed firms have focused on developing novel ways to apply information technology and the widespread diffusion of mobile communications. One manifestation has been platforms that connect and employ widely dispersed sellers of services and goods (frequently dubbed the "sharing economy," and manifested by companies such as Airbnb and Uber). A second has been firms that substantially improve the efficiency of existing services at much lower price points: for example, the ways in which Salesforce.com and other companies provide "Software as a Service" to businesses, the rise of "fintech," and the plethora of "Mobile Apps" available for consumers. Third, several companies replicated business models successful in the United States in other national markets, with the Chinese companies Alibaba and Tencent being the most dramatic exemplars. An important

consequence was the dramatic increase in venture capital investment in Asia over this period (a topic worthy of much closer study).

Another demand driver has been the substantially lower cost of starting a new business in the software and services sectors. As documented by Ewens, Nanda, and Rhodes-Kropf (2018), the much lower initial capital needed for new ventures in these sectors made it much cheaper to learn about their potential. This change led early-stage investors to be more willing to fund less proven (but potentially high-return) ideas and entrepreneurial teams in these sectors. One manifestation was an increase in a "spray and pray" investment approach, where financiers provide a small amount of funding and limited governance to a larger number of startups. As seen in Figure 1, the number of startups receiving a first round of venture capital financing rose substantially over this period.

Coinciding with the fall in cost of starting businesses and the entry of less-experienced founding teams has been the emergence of complementary institutions that fund and mentor very early-stage entrepreneurs. For example, the substantially smaller quantum of capital required to get a business off the ground has led to more opportunities for angel (or individual) investors. Not only did angel groups grow in size (Kerr, Lerner, and Schoar 2014), but some angel investors (the "Super Angels") even began to raise small funds to finance startups at earlier stages than was typical for institutional venture capital investors. Further, using online platforms such as AngelList, groups of individuals could back a lead investor who aimed to replicate some of the systematic diligence and monitoring functions played by traditional venture groups (Agrawal, Catalini, and Goldfarb 2016). The contemporaneous rise of equity crowdfunding and initial coin offerings has had a more mixed legacy, enabling widespread participation in financing startups by the populace but also raising concerns about fraud (Howell, Niessner and Yermack forthcoming; Lin 2017; Zetzsche et al. 2018).[3] The evolving early-stage market also created an increasingly important role for business accelerators, which sought to systemize the mentoring and development of the larger number of inexperienced, first-time entrepreneurs receiving financing (Gonzales-Uribe and Leatherbee 2017; Hochberg 2016).

The last decade has also seen substantial changes in the way that venture capital-backed firms grow and achieve exits for their investments. One element of this shift is the marked decline in the number of initial public offerings since the "dot com" bust in 2000. Instead, venture capitalists are far more likely to exit investments through acquisitions. Inasmuch as firms are going public, they are doing so at more mature stages in their life-cycle (Ewens and Farre-Mensa 2020).

Understanding the drivers for such shifts—and the more general reduction in the number of publicly traded US firms—is challenging. Potential explanations include technological shifts leading to a rise in platform (winner-take-all) businesses (Gao, Ritter, and Zhu 2013), regulations making it harder for small firms to

[3] A parallel literature has examined the rise of reward-based crowdfunding (Mollick 2014) and peer-to-peer *lending* platforms (Iyer et al. 2016) which we do not discuss here due to the focus on equity finance of technology-based ventures.

go public (Iliev 2010), changes in securities laws that facilitated the flow of more capital into private markets (Ewens and Farre-Mensa 2020), and monetary policy following the financial crisis.

But whatever the cause, the fact that firms that are more mature when they go public has also meant that they do so at substantially higher valuations. Investors that traditionally focused solely on the public markets saw that they were missing out on the capital gains that companies such as Facebook, LinkedIn, and Salesforce garnered while still private. These investors consequentially sought out opportunities in the private venture capital market.

As a result of this interest, the past decade saw an increase in the number of venture capital funds raising capital. The most conspicuous impact of this flood of capital into venture capital was the rise of "mega-funds," which refers to venture capital funds that are substantially larger than historical averages. The most salient of these was SoftBank's Vision Fund. At the time of its first closing on $93 billion in May 2017 (with an anchor investment of $45 billion from the Saudi Public Investment Fund), it was already 30 times larger than the previous largest venture capital fund raised (New Enterprise Associates' 2015 Fund XV). SoftBank would ultimately go on to raise $100 billion for its fund. This rapid increase in interest in venture capital also triggered traditional venture firms to raise very large funds, such as the $8 billion Sequoia Capital Global Growth Fund III in 2018.

In addition, frustration with the high fees charged by venture capital firms led sovereign wealth funds, hedge funds, mutual funds, and other public market investors to begin making direct investments into firms backed by venture capital (Fang, Ivashina, and Lerner 2015; Lerner et al. 2018). Ewens and Farre-Mensa (2020) estimate that between 2014 and 2016, over three-quarters of the late-stage venture capital funding came from such non-traditional investors. Whether through large funds or direct investments, much of the capital from these later-stage investors has gone to "unicorns," defined as privately held firms with nominal valuations in excess of $1 billion.

This combination of new entrants deploying small amounts of capital at the early stage with the rise of mega-funds is reflected in fund size statistics. These changes are best illustrated by looking "peak to peak," from 2007 to 2019 in Table 2. The size of the median fund raised by venture capital investors has fallen from $133 million to $80 million. Meanwhile, the number of funds with $1 billion or more of capital rose from three in 2007 to eight in 2019 (NVCA 2020).

Venture Capital's Limitations

The growth of the venture capital market in the past decade should not blind us to its limitations as an engine of innovation. Indeed, the changes delineated in the previous section will likely exacerbate these challenges. We lay out three distinct areas of concern about venture capital and its ability to successfully spur innovations. While the discussion must be inherently more speculative, given the relatively

Table 2
The State of US Venture Capital Funds in 2007 versus 2019

	2007	*2019*
Number of firms that raised funds in the prior 8 years	946	1,328
Number of VC funds raising money in that year	187	272
Number of funds greater than $1 billion in size	3	8
VC capital raised (billions of $)	35	51
Total VC AUM (billions of $)	222	444
Median fund size (millions of $)	133	80
Average fund size (millions of $)	213	189

Source: Data are drawn from the National Venture Capital Association's 2020 Yearbook and accompanying supplemental data pack. Data for "number of funds greater than $1 billion in size" are drawn from the PitchBook database.
Note: The NVCA includes in its fundraising data "only funds based in the United States that have held their final close," while its deals data include financings of companies headquartered in the United States but potentially from investors based outside the United States.

limited work done in this area, we suggest that these questions would benefit from scholarly attention in the years to come.

Optimized for a Narrow Slice of Technological Innovation

Despite the substantial growth of venture capital in the four decades since the revision of the "prudent man" rule for pension fund investment in 1979, venture capital touches only a tiny share of firms in the American economy. The estimated $450 billion currently under management by US venture capital firms (NVCA 2020) remains small in comparison to the several trillion dollars managed by the broader asset class of all US private equity, not to mention the total of all US public equities, estimated at $42.9 trillion at the end of 2018 (SIFMA 2019). Only a few thousand new firms each year raise institutional venture capital for the first time, as compared to over 600,000 annual business starts in the United States. Even among high-potential firms engaged in innovation, Farre-Mensa, Hegde, and Ljungqvist (2020) found that only 7 percent of firms that filed for a patent went on to raise institutional venture capital. These disparities are likely even more extreme in other nations, where the venture industry is less mature.

One reason for this is structural. Venture capital investors typically raise funds for a specific (usually a ten-year) period. This time frame implies that venture capitalists are naturally drawn to investment opportunities where the ideas can be commercialized and their value realized through an "exit" within a reasonably short period. Sudden market downturns, as occurred in 2000, 2008, and 2020, may disrupt plans to exit investments, creating more pressure to sell when market conditions permit, even if earlier than optimal for the firm. These constraints imply that venture capital investors often exit their investments well before growth opportunities are fully realized. As a result, they are often drawn to sectors with large uncertainty about an idea's potential that can nevertheless be resolved quickly.

What leads to variation in the degree to which uncertainty about the prospects of a young firm can be reduced quickly? An important element appears to be the

nature of uncertainty about demand for the new product or service. Put another way, can uncertainty about the viability of the offering and the market demand be resolved quickly? Software and service businesses—which are typically based on proven technologies, often have short development times and can benefit from quick market feedback—are amenable to this approach. Also, as noted above, technological changes over the past two decades have made it quicker and cheaper to learn about demand for a new software business. By way of contrast, many other sectors like clean energy, new materials, and others are less amenable to such rapid learning. The widespread interest among venture capital investors in the few exceptions, such as biotechnology startups, is tied to the drug approval and reimbursement system that enables investors to project the market value of a new drug accurately if it is successful in passing through clinical trials (Janeway 2018).

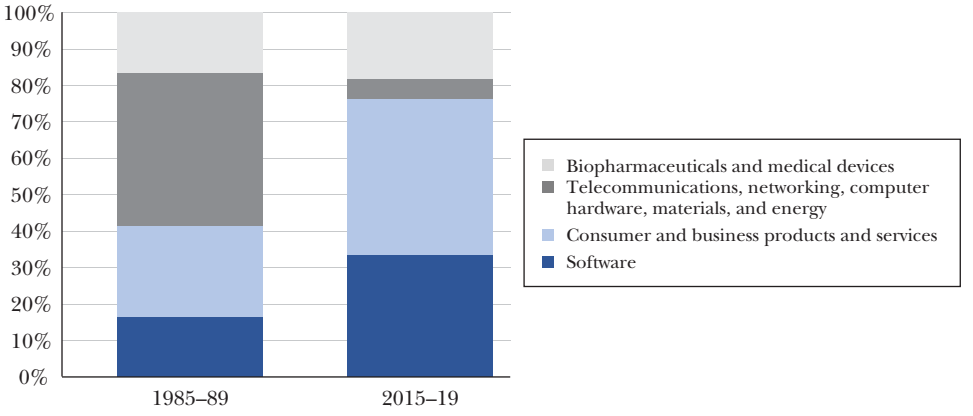
This suggestion is underscored by computations by Sand Hill Econometrics. This firm creates a series of indexes (described in Hall and Woodward 2004) that seek to capture the gross returns (that is, before management fees and profit-sharing) from investments in all active venture transactions in a given category. Their calculations suggest that an investment in the software deals between December 1991 and September 2019 would have yielded an annualized gross return of 24 percent per annum, far greater than investments in hardware (17 percent), healthcare (13 percent), or clean-tech (2 percent). This index further shows that the divergence in the performance of these categories has been particularly stark in the last decade.

These stark differences in economics have unsurprisingly led to shifts in the composition of venture capital portfolios. Examining the portfolio of a single venture group four decades apart demonstrates the extent of this focus and narrowing.[4] Charles River Ventures was founded by three seasoned executives from the operating and investment worlds in 1970. Within its first four years, it had almost completely invested its nearly $6 million first fund into 18 firms. These included classes of technologies that would be comfortably at home in a typical venture capitalist's portfolio today: a startup designing computer systems for hospitals (Health Data Corporation), a software company developing automated credit scoring systems (American Management Systems), and a firm seeking to develop an electric car (Electromotion, which, alas, proved to be a few decades before its time). Other companies, however, were much more unusual by today's venture standards: for instance, startups seeking to provide birth control for dogs (Agrophysics), high-strength fabrics for balloons and other demanding applications (N.F. Doweave), and turnkey systems for pig farming (International Farm Systems). Only eight of the 18 initial portfolio companies—less than half—were related to communications, information technology, or human health care.

The portfolio of Charles River Ventures looks very different in December 2019. Of the firms listed as investments, about 90 percent are classified as being related

[4]This example is drawn from Banks and Liles (1975) and "Charles River Ventures," http://www.crv.com/.

Figure 2
Venture Capital Investment into US Startups between 1985 and 2019, by Sector



Source: Data are drawn from the National Venture Capital Association's yearbooks and related resources. Software refers to firms classified as being in the Software industry. Consumer and Business Products and Services refer to startups in the following categories: Business Products and Services, Consumer Products and Services, Financial Services, Healthcare Services, IT Services, Media and Entertainment and Retailing/Distribution. Telecommunications, Networking, Computer Hardware and Energy refer to startups in the following categories: Computers and Peripherals, Electronics/Instrumentation, Networking and Equipment, Semiconductors, Telecommunications, Industrial/Energy and Other. Biopharmaceuticals and Medical Devices refer to startups in the following categories: Biopharmaceuticals and Medical Devices and Equipment
Note: This exhibit reports investment by venture capital investors into US startups between 1985 and 2019, broken down by four distinct sectors.

to information technology comprising social networks, applications for consumers, and software and services related to enhancing business productivity. Approximately 5 percent of investments are classified as being related to health care, materials, and energy.

This shift in Charles River's portfolio reflects the patterns of the industry at large, as Figure 2 depicts. By way of preface, it should be noted that it is difficult to trace fine-grained industry categories over multiple decades. New categories of firms (such as social networks and digital media) have emerged. Moreover, firms do not always fit neatly into a single classification: for example, Uber is a software firm, a transportation firm, and a consumer service firm. With these caveats duly noted, in Figure 2 we categorize firms into four broad classifications that are reasonably comparable over time: computer software, hardware of many types (from energy to instruments to semiconductors), business and consumer products/services, and medical, including biopharmaceuticals.

Hardware dominated the investments made by venture capital in the period from 1985 to 1989, accounting for 42 percent of dollars invested into startups by venture capital. Software and service startups accounted for nearly the same share of investments made in 1985–1989, while biopharmaceuticals and medical devices represented most of the remainder. The figure also shows the large shift in focus

of venture capital firms away from hardware and towards software and service businesses. Biopharmaceutical and medical device startups have received approximately the same share of funding. However, the chart masks the fact that this investment now comes from a smaller share of venture capital firms specializing in this sector, as opposed to from many generalist venture capital firms.

The same concentration of investment on software and related businesses seen in Figure 2 is also seen in the patent data. Using data on the patents filed at the US Patent and Trademark Office over the period 2008–2017, we found that the top ten patent classes using the US Cooperative Patent Classification (CPC) system represented 48 percent of all venture capital patents filed over the 2008–2017 period, as compared to 24 percent for the top ten patent classes for patents not filed by venture capital–backed firms.

Thus, while venture funding is very efficacious in stimulating a certain kind of innovative business, the scope is increasingly limited. This concentration may be privately optimal from the perspective of the venture funds and those who provide them with capital. It is natural to worry, however, about the social implications of these shifts. For instance, promising startups developing renewable energy technologies and advanced materials, which might have broad societal benefits, may languish unfunded.

The reader might well raise eyebrows at this suggestion. If value-creating entrepreneurial investment opportunities exist, should some other investor step in? Certainly, we have seen corporations in sectors such as energy making investments in young firms. But as we will discuss in the final section, these efforts to date have been far from resounding successes.

The Disproportionate Role of a Few Deep-Pocketed Investors

Venture activity is concentrated. Yes, the National Venture Capital Association estimates that there were a little over 1,000 US venture capital funds in 2019. But a small number of large venture capital firms hold the vast majority of capital.

To illustrate this point, we created a list of all institutional venture capital investors that made at least one investment into a US-headquartered startup in 2018. For these investors, we examined the total funds they had raised from 2014 to 2018: approximately $284 billion raised by 985 investors. Looking at the concentration in the capital raised by these investors provides a good proxy for the concentration in assets under management across institutional venture capital investors. The top 50 investors, or about 5 percent of the venture capital firms, raised half of the total capital over this period.[5]

The large inflow of capital in the last decade has further concentrated capital in the hands of top funds. The reasons for this are worth further inquiry but are likely, at least in part, due to strong persistence in the relative performance of the venture

[5] Taking into account non-traditional investors such as SoftBank increases the number of investors under consideration to 1,074. Among this larger set of investors, the top 50 investors accounted for 68 percent of the total capital raised by investors over this period.

capital funds (Kaplan and Schoar 2005; Harris et al. 2014).[6] A small number of higher-performing venture capital firms have continued to raise ever-larger funds.

Moreover, these deep-pocketed investors can play a disproportionate role in driving where other investors put their money. Investors with smaller sums of capital under management typically focus on investments at the earliest stages of a startup's life, well before the startup is profitable or even has revenue. These early-stage venture capital investors often do not have the capital to continue financing startups across subsequent rounds to the point where the firm can be sold for an attractive valuation. Thus, they are dependent on their larger peers to step in and continue financing the firms they initially funded. Consequentially, a major worry for early investors is that an otherwise healthy startup might not be able to raise follow-up capital (Nanda and Rhodes-Kropf 2017). In this way, the preferences of large late-stage investors can shape where early-stage investors are willing to invest. Consistent with this suggestion, Howell et al. (2020) find that early-stage venture capital appears to be particularly sensitive to market conditions when examining recessions during the past half-century.

Concerns about a small number of financiers acting as gatekeepers may be particularly salient when considering the characteristics of these financiers. We highlight three dimensions. First, major venture funds are based in a handful of places. In the United States, National Venture Capital Association statistics suggest that three metropolitan areas—the San Francisco Bay Area, Greater New York, and Greater Boston—account for about two-thirds of the venture capital deployed by firms each year.[7] The same phenomenon also seems to manifest itself globally, though good statistics are hard to find. For instance, a tabulation of PitchBook data between 2015 and 2017 by Florida and Hathaway (2018) concludes that the top 25 urban areas accounted for 75 percent of all disbursements globally. Given this concentration of capital, the startup community has rearranged itself to "follow the money."

Why might the resulting geographic concentration be a cause for concern? After all, economists have long pointed out that there are increasing returns to scale in entrepreneurial and innovative activity. Regions like Silicon Valley have an abundance of resources for entrepreneurs, ranging from excellent engineers used to working long hours for risky stock options, knowledgeable patent attorneys, and of course, lots of financiers. As a result, there are real social benefits from geographic concentration of entrepreneurs.

On the other hand, the geographic concentration of venture capital has probably accelerated the "hollowing out" of innovative activities in many other parts

[6]Related work has examined the drivers of this persistence and the degree to which it might be a consequence of differences in skill across venture capital investors versus other factors such as sorting or preferential access to deal flow that may perpetuate initial differences in performance across investors (Sørensen 2007; Hochberg, Ljungqvist, and Lu 2007; Korteweg and Sørensen 2017; Ewens and Rhodes-Kropf 2015; Nanda, Samila, and Sorenson 2020).

[7]For details, see the National Venture Capital Association Yearbook at https://nvca.org/wpcontent/uploads/2019/08/NVCA-2019-Yearbook.pdf and earlier editions archived at https://nvca.org/pressreleases_category/research/.

Table 3
Characteristics of Key US-based Investment Professionals in the 50 Largest Venture Capital Firms

	US-based partners	*US-based partners with at least one board seat*
Total number of Partners	416	265
Share male	82%	92%
Share attended top universities	59%	72%
Share with MBA from Harvard	12%	15%
Share with MBA from Stanford	9%	13%
Share located in Bay Area	69%	69%
Share located in Greater Boston	9%	11%
Share located in New York City	14%	11%
Average number of board seats held		6.1
Median number of board seats held		5

Source: Data are drawn from the PitchBook database. We first restrict investment professionals in these firms to titles that are one of Managing General Partner, Managing Partner, Founding Partner, General Partner, Senior Partner, or Partner and further restrict them to individuals based in the United States. In column 2, we examine a subset of these individuals who also sit on at least one board seat, as some of the firms in our sample have a larger number of individuals with a "Partner" title than those who make investment decisions or are actively involved in governing startups.
Note: This table reports characteristics of the key US-based investment professionals working for the 50 largest venture capital firms reported in Exhibit 5.

of the world. Venture firms based in other cities might have chosen very different firms to invest in given their perspectives on their local economies. More generally, Glaeser and Hausman (2020) have documented in the United States the growing hubs of innovative activity in places far removed from the areas with the greatest economic need, a phenomenon that the growth of venture capital has accelerated.

Second, the background of individual decision-makers at venture firms remains far from representative of the general population. Table 3 focuses on the 417 individuals listed as US-based partners of the top 5 percent of venture capital firms noted above.[8] Some firms give "partner" titles to a larger number of individuals than the true decision-makers in the partnership, so we also examine the subset of these individuals who sit on at least one corporate board. This restriction narrows the set of individuals to 265.

Table 3 documents the composition of this group. Eighty percent of partners are male; among the set of partners with at least one board seat, 91 percent are male. Three-quarters of partners with at least one board seat attended either an Ivy League school, or one of Caltech, MIT, or Stanford; moreover, nearly 30 percent

[8]We used data from the PitchBook database (https://pitchbook.com/) for this analysis. We restricted the investment professionals listed for each of these firms to those with the titles of Managing General Partner, Managing Partner, Founding Partner, General Partner, Senior Partner, or Partner. We further limited them to individuals based in the United States.

of these individuals are graduates of just Harvard Business School or the Stanford Graduate School of Business. In terms of location, 69 percent are based in the Bay Area alone and over 90 percent are based in either the Bay Area, Greater Boston, or New York.

The nonrepresentative nature of the decision makers at these firms is important because of the growing evidence that a lack of diversity among venture capitalists has an impact on what businesses get funded. For example, Gompers and Wang (2017) use the number of daughters of venture capital partners as an instrumental variable and show that it is correlated with a higher proportion of female partners and improved deal and fund performance. Ewens and Townsend (2020) document that male and female investors appear to have gendered preferences (or respond to different signals about potential cash flows) in terms of the companies they back. Understanding whether such frictions are consequential enough to influence the nature of innovations that are backed and the choice of products faced by consumers is an important question that we believe deserves more research.

A final concern, more difficult to document, relates to the criteria that these investors use to make decisions more generally (Gompers et al. 2020). While academics have spent a great deal of time seeking to understand the structure of venture investment agreements and post-deal involvement, the process before the transaction is much less understood. We understand that early-stage investors rely heavily on signals of entrepreneur quality (Bernstein, Korteweg, and Laws 2017), but know very little as to whether the emphasis on these signals is efficient. Recent work by Cao (2019), for example, shows that information frictions from early-stage platforms can lead to systematic downstream effects on firm funding. Given the increasing importance of venture capital for innovation and growth, understanding the way in which these investors acquire and aggregate signals of a venture's potential and the frictions in this process are important and promising areas of future research.

A Declining Emphasis on Governance?

The third concern we highlight here has to do with the seeming decline in active corporate governance by venture capital funds. Venture capital has traditionally been a tough business, with onerous agreements in which firm founders gave venture capital firms significant stock ownership in exchange for funding (Kaplan and Strömberg 2003). Moreover, this stock ownership was not just "paper rights": frequent turnover of management driven by venture capital was traditionally the rule (Kaplan, Sensoy, and Strömberg 2009; Ewens and Marx 2018). These patterns have changed dramatically in the past decade. Across the board, "founder friendly" terms appear to have replaced the more onerous provisions traditionally demanded by venture capitalists.

Several potential explanations can be offered for this change. First, the intense competition between venture capitalists during the 2010s may have led to better terms for corporate founders. Even in less ebullient markets, the most promising entrepreneurs have a lot of discretion from whom they choose to receive funding (Hsu 2004): venture returns are very skewed, with a few deals generating the bulk

of the returns (Hall and Woodward 2010). This pattern has been especially true in the last few years, given the proliferation of mega-funds and the explosion in capital from non-traditional investors such as SoftBank, sovereign wealth funds, and corporations. In an intensely competitive market, some venture capital firms may be tempted to pitch entrepreneur-friendly contracts to founders in an attempt to get access to the most attractive deals (Eldar, Hochberg, and Litov 2020). Reflecting this competition, venture capital groups may have chosen to outdo each other in the extent of their hospitality toward company founders.

To an economist, however, this explanation raises new puzzles. If the intensive governance provided by venture capitalists is socially beneficial—as generations of academic analyses would suggest—why would groups choose to abandon it? Should not venture firms compete instead by offering entrepreneurs progressively higher valuations (and less dilution of their initial equity stakes), not by abandoning governance provisions? Does this explanation also imply that firm founders may underestimate the need for governance?

Other possible explanations for the decline in governance, however, may suggest deeper structural drivers of this trend. For example, a possible explanation reflects the changing dynamics of early- and later-stage investing discussed above. It has become far cheaper to start a new business. Perhaps the capital that firms need at the earliest stage is too small to make it worthwhile for venture capital firms to engage in active governance. Indeed, some venture capital firms have adopted the "spray and pray" investment strategy at the seed stage of financing, in which they focus on learning about the potential of a venture before spending time governing it (Ewens, Nanda, and Rhodes-Kropf 2018).

In addition, the massive inflow of venture capital from investors that usually focus on the public market may have changed the focus of contractual rights at later stages. A single fund manager in these entities may hold hundreds of separate firms and have little experience directly governing the firms in their portfolio. These passive investors are unlikely to have the capabilities to provide effective governance to the entrepreneurial ventures. As Chernenko, Lerner, and Zeng (2017) document, mutual funds seem far more concerned with ensuring that there is a path to liquidity, reflecting the short-term nature of the capital that they have raised from investors. The changing composition of the capital sources may have thus also led to a reduced focus on governance.

There is a reason to fear that even among traditional venture capital investors, governance may decline for structural reasons. As firms stay private longer, venture capital investors may end up sitting on a larger number of company boards. Put another way, the classic structure of venture funds may have begun to get overwhelmed by the flow of outside money, new financial intermediaries, and the associated change in practices. As a result, venture capital-backed firms may not be receiving the same degree of governance.

Whatever the causes, there are a number of high-profile examples in recent years in which the charismatic founder of a "unicorn" company has been ousted. The departures of founders and chief executive officers like Travis Kalanick at Uber,

Elizabeth Holmes at Theranos, and Adam Neumann at WeWork are quite different in their details. But overall, they illustrate some consequences of allowing entrepreneurs with limited prior management experience to raise enormous sums for new ventures with little in the way of formal oversight and governance protections. Understanding why traditional venture capital contractual provisions have faded in importance and their social welfare implications appears to be a promising area of future research for both theorists and empiricists alike.

New Approaches for Venture Capital

Venture capital has been a highly efficacious way to support certain kinds of innovation, as reflected in the importance of venture-backed companies in the stock market and in the economy as a whole. At the same time, the industry has important limitations. We now turn to some ideas that might help the venture capital industry to become more effective, with a recognition that these hypotheses may be of greater interest to those practitioners and academics interested in thinking "outside the box." We highlight two sets of ideas: the first, which owes a heavy debt to Lerner (2012), relates to the organizational and incentive structure of venture capital partnerships; the second, to the way in which venture capital firms focus on managing their investments in more challenging sectors.

Rethinking Venture Capital Partnerships

Since the early days of the industry, venture capital funds have been eight to ten years in length, with provisions for one or more one- to two-year extensions. Venture capitalists typically have five years in which to invest the capital and then are expected to use the remaining period to harvest their investments.

The uniformity of these rules is puzzling. Funds differ tremendously in their investment foci: from quick-hit social media businesses to long-gestating biotechnology projects. In periods when the public markets are enthusiastic, venture capitalists may be able to exit still-immature firms that have yet to show profits and, in some cases, before they even have revenues. But as discussed above, there is tremendous variation in the maturation of firms in different industries. Certainly, within corporate research laboratories, great diversity across industries exists in terms of the typical project length. What explains the constancy of the venture fund lives?

One possible explanation is that a reasonably short fund life seems to have been the norm in limited partnerships of all types. For example, many of the other arenas where limited partnerships were employed in the 20th century, such as real estate, oil-and-gas exploration, and maritime shipping, all were reasonably short-lived. In the formative days of venture partnerships, the lawyers drafting the agreements may have gravitated to the relatively short fund lives that were common in other contexts. With the passage of time, such arrangements have then been taken as gospel by limited and general partners alike.

Another factor behind the persistence of the ten-year agreement has been the resistance of limited partners—that is, the investors in the venture capital funds—to longer fund lives. These investors may fear that if they give the funds to a sub-standard venture group for a longer period, they will be stuck paying fees until the end of time for very limited returns. This reluctance may tell us more about the outsized nature of the fees that venture funds receive than about the inherent desirability of a longer-lived fund.

Indeed, the manner in which venture capitalists are compensated has changed little, even as the funds have grown much larger. Venture groups typically receive a share of the capital gains they generate (typically 20 percent, but sometimes as high as 30 percent) and then an annual management fee (often between 1.5 percent and 2.5 percent of capital under management, though it often scales down in later years). Such fees are quite modest for a fund of only a few million dollars in size: it is likely to cover only a very modest salary for the partners once the costs of an office, travel, and support staff are factored in.

But this compensation structure has remained largely unchanged as funds have become substantially larger. Moreover, as venture capital groups begin managing hundreds of millions or billions of dollars, substantial "economies of scale" appear: put another way, as a group becomes ten times larger, expenses increase much less than tenfold. As a result, management fees themselves become a profit center for the firm. These steady profits may create incentives of their own which may not be very appealing to investors. For instance, there will be an incentive to raise a larger fund at the expense of lower returns, which in turn may be tied to the greater concentration of capital held by a few investors; an incentive to put funds to work quickly and with a subpar amount of vetting so that a new fund can be raised sooner; and an incentive to focus on excessively safe investments that will not have as much upside but will pose less risk of a franchise-damaging visible failure.

Just how large a temptation the venture capital compensation scheme can pose is illustrated in the work of Metrick and Yasuda (2010), who show that of every $100 invested by the limited partners, over $23 end up in the pockets of the venture investors. These sums might not be disturbing if the very substantial payouts to each partner reflected even larger returns being made by the limited partners in the fund. But profit sharing is not the most important source of compensation. Instead, almost two-thirds of the income (in time-adjusted dollars) is coming from the venture capital management fees, which remain fixed whether the fund does well or poorly. These incentives clearly may motivate groups to add capital in excess of the growth of partners, even if performance suffers somewhat.

Interestingly, an alternative model already exists: the way that venture capital groups used to operate. Early venture capital groups, beginning with the pioneering partnership of Draper, Gaither & Anderson, negotiated budgets annually with their investors. The venture capitalists would lay out the projected expenses and salaries and reach a mutual agreement with the limited partners about these costs. The fees would be intended to cover these costs, but no more. (A few "old school" groups such as Greylock still use such an arrangement.) Such negotiated fees greatly reduce

the temptation to grow at the expense of performance and ultimately are likely to lead to more successful and innovative startups.

What explains the traditional reluctance of the limited partners who invest in venture capital funds to push to change these compensation arrangements? Staff members may not really understand the economics of the funds, or they may fear that rocking the boat would limit their own ability to get a high-paying position at a fund or an intermediary in the future. Alternatively, the officers may worry that developing a reputation as an activist would jeopardize their organization's ability to access the funds with the highest returns. The last concern is a reasonable one. After the giant California Public Employees' Retirement System led a consortium of pension funds who pushed for an overhaul of private equity compensation in the mid-1990s, they were shunned by venture and buyout funds alike. In recent years, we have seen more collective discussion of these issues by limited partners in meetings of the Institutional Limited Partners Association. But many of their proposals have been modest half-measures, without addressing the more fundamental issues.

Yet another question is why such innovations are not adopted by newer venture firms as a way to differentiate themselves? It might be thought that such "LP-friendly" terms might attract new investors. While examples along these lines have occurred, anecdotal evidence suggests that many groups who have tried such an approach have found it to be tough sledding. In part, limited partners may interpret such concessions as an adverse signal, indicative of a lack of confidence on the part of the new fund managers in their ability to raise a fund. Additionally, even if the group responsible for private equity investments at the pension fund understands the advantages of the proposed alternative arrangement, the investment committee that ultimately makes the decisions at the pension fund may not. These information problems may lead to the persistence of a socially suboptimal fund structure.

New Approaches to Managing Investments

Venture capital firms face huge uncertainty about the ultimate potential of startup firms. Indeed, over half the investments of even the most successful venture capital investors fail, while the vast majority of returns are generated by a few extremely successful investments that are hard to predict upfront.

As noted before, venture capital investors are drawn to sectors in which this uncertainty can be reduced quickly. Staged financing is therefore not only valuable to venture capital investors as a governance tool, but also as a method of learning about the startup's potential through a sequence of investments over time. This approach to evaluating and governing ventures in software industries has been outlined in practitioner guides such as *The Lean Startup* (Ries 2011).

But the ability to learn quickly about the promise of new ventures is harder if there is substantial regulatory, technology, and market risk, which we suggest explains the poor performance and declining share of venture investment outside of software. For example, forecasting the unit costs associated with energy storage at scale using a new battery material can be extremely difficult, even if the technology works in a controlled laboratory environment. Because uncertainty about market

demand is tied to firms' ability to produce at a certain price point, forecasting demand in this setting is hard.

One promising response to these challenges is to rethink the organizational model for incubating and financing "tough tech" ventures. The venture approach entails entrepreneurs coming to venture capital firms to pitch them new ideas and the firms deciding whether to fund them. This approach has the benefit of enabling the investors to maintain an arm's length relationship from the entrepreneurial team, reducing the entrenchment that is sometimes associated with corporate research and development and internal capital markets.

An alternative approach that has begun to be used by some venture capital investors specializing in biopharmaceuticals (such as Third Rock Ventures and Flagship Pioneering)is to incubate and finance ideas in-house. This process has the benefit of reducing asymmetric information because much of the staff for the team of entrepreneurs comes from within the fund. It also enables the venture capital firm to fund what it might believe is the most promising idea or approach as opposed to selecting among the ideas that walked in the door. A related approach is illustrated by Breakthrough Energy Ventures, which has a team of in-house scientists who jointly make investment decisions with traditional investment partners. Such new approaches may hold promise for widening the scope of venture capital investment. Understanding the tradeoffs associated with bringing incubation inside the venture capital firm and organizing new ventures more like corporate research and development seems to be a promising area of academic research.

It is also natural to wonder whether collaboration with other parties—governments, non-profits, and corporations—might alleviate some of the barriers to financing new ventures in more difficult technologies. Of course, this suggestion is not new. Governments have been involved with the promotion of venture capital at least since 1946, when a consortium of the Bank of England and leading British banks combined to create the British firm 3i as a vehicle to make long-term investments in smaller firms. Corporations have been collaborating with venture capitalists since the 1960s. Universities and other nonprofits have been incubating, mentoring, and directly financing new ventures for much of the last half-century.

But the track record of these collaborative efforts has been quite mixed. There have been successes, such as the Israeli government's jump-starting of its venture industry through the Yozma program that leveraged public money to attract private investment, or the success of many pharmaceutical firms in responding to the biotechnology revolution through their venturing initiatives. At the same time, anecdotes abound of naïve officials making poor decisions. For instance, the leadership of Boston University put one-third of the university endowment into a single faculty-founded biotechnology company, Seragen, an investment that was ultimately sold for pennies on the dollar.[9]

[9]This account is drawn from Seragen's filings with the US Securities and Exchange Commission; the annual reports of the National Association of College and University Business Officers; the reporting of

The statistical evidence on collaborative efforts, while limited, does not seem inspiring. For instance, the Thomson Reuters (now Refinitiv) database suggests that between 1993 and 2013, corporate venture funds lost 4 percent per annum, at a time when US venture funds had annual returns of nearly 30 percent (for a more optimistic view, see Ma 2020).

We believe that these collaborations can be beneficial, but only if executed correctly. This caveat is important. As an example, we will highlight the role of US government in the venture market. The primary mechanism through which government policy interacts directly with new ventures is through the Small Business Innovation Research (SBIR) program. A striking study by Howell (2017) suggests that the initial Phase I awards under this program have very positive effects on new technology ventures, doubling the probability that a firm receives venture capital and boosting patenting and revenue. But these Phase I awards made up only 20 percent total of the $2.8 billion spent on the program (US Small Business Administration 2018). The bulk of the funding goes to larger Phase II awards, which Howell argues have no positive impacts. Similarly, both Howell (2017) and Lerner (1999) document that a relatively small number of companies capture a disproportionate number of awards. These "SBIR mills" commercialize far fewer projects than those firms that receive just one or a handful of SBIR grants, but the repeat winners often have active staffs of lobbyists in Washington scouring for award opportunities. Despite these well-understood issues, the design of the program looks virtually identical to its initial manifestation in 1977. There has been almost no serious discussion in Washington regarding the idea of shifting more SBIR resources to Phase I grants or curtailing grants to "mills." The experience of the SBIR program underscores the need for careful initial design, painstaking evaluation, and a willingness to redesign initiatives.

One promising area of recent growth has been the interest among philanthropic organizations in financing early-stage, high-risk research and development. The hope is that once sufficient development of the idea has taken place, private venture capitalists will be willing to step in. Such activities have been most visible of late in the early-stage financing of vaccines, including for COVID-19. Beyond healthcare, efforts are also emerging to finance initial investments in sectors that have substantial potential societal benefits but large risks. Illustrations include the Prime Coalition's funding of companies that combat climate change yet are sufficiently risky to deter traditional investors (at https://primecoalition.org/what-is-prime/) and the initiative from the Ford and Rockefeller Foundations to seed venture funds investing in regions that have traditionally not attracted such capital (as reported in Murray 2020; see also the Community Development Venture Capital Alliance at https://cdvca.org/about-us/missionhistory). While these efforts are likely to face substantial challenges, they also have real potential.

Barboza (1998); and the decision of the Court of Chancery of Delaware in *Oliver v. Boston University*, C.A. no. 16570-NC. (Del. Ch. Apr. 14, 2006).

Final Thoughts

The growth of venture capital in the past decade, both in the United States and worldwide, is an important validation of the underlying model. At the same time, the period has brought into sharp relief the structural challenges facing the industry.

Over the past decades, academics and practitioners alike have highlighted the strengths of venture capital. Understanding and articulating its limitations as well as how institutional innovations can address them, is an important challenge to both groups going forward.

■ *Harvard Business School's Division of Research provided funding for this work. Terrence Shu provided excellent research assistance. The ideas in this essay draw, among other sources, on those in Gompers and Lerner (2001a); Kerr, Nanda, and Rhodes-Kropf (2014); Lerner (2012); and Ivashina and Lerner (2019). We thank Gordon Hanson, Enrico Moretti, Timothy Taylor, and Heidi Williams for valuable feedback. We owe a debt of gratitude to Paul Gompers, Bill Janeway, Steve Kaplan, Victoria Ivashina, Matthew Rhodes-Kropf, William Sahlman, and especially Felda Hardymon for many helpful conversations over the years. We thank Jeremy Greenwood for pointing out the Arrow interview. Lerner has received compensation from advising institutional investors in venture capital funds, venture capital groups, and governments designing policies relevant to venture capital.*

References

Agrawal, Ajay, Christian Catalini, and Avi Goldfarb. 2016. "Are Syndicates the Killer App of Equity Crowdfunding?" *California Management Review* 58 (2): 111–124.

Ante, Spencer E. 2008. *Creative Capital: Georges Doriot and the Birth of Venture Capital.* Boston: Harvard Business School Press.

Arrow, Kenneth. 1995. "Interview with Kenneth Arrow, Federal Reserve Bank of Minneapolis." https://www.minneapolisfed.org/article/1995/interview-with-kenneth-arrow.

Astuti, Guido. 1933. *Origini e Svolgimento Storico della Commenda Fino al Secolo XIII.* Milan: S. Lattes & Co.

Banks, Robert L., and Patrick R. Liles. 1975. "The Charles River Partnership." Harvard Business School Case 375075.

Barboza, David. 1998. "Loving a Stock, Not Wisely but Too Well." *New York Times*, September 20. http://www.nytimes.com/1998/09/20/business/loving-a-stock-not-wisely-but-too-well.html.

Bernstein, Shai, Arthur Korteweg, and Kevin Laws. 2017. "Attracting Early Stage Investors: Evidence from a Randomized Field Experiment." *Journal of Finance* 72 (2): 509–38.

Bernstein, Shai, Xavier Giroud, and Richard R. Townsend. 2016. "The Impact of Venture Capital Monitoring." *Journal of Finance* 71 (4): 1591–1622.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb. 2020. "Are Ideas Getting Harder to Find?" *American Economic Review* 110 (4): 1104–44.

Cao, Ruiqing. 2019. "Crowd-Based Rankings and Frictions in New Venture Finance." Unpublished.
Chemmanur, Thomas J., Karthik Krishnan, and Debarshi K. Nandy. 2011. "How Does Venture Capital Financing Improve Efficiency in Private Firms? A Look beneath the Surface." *The Review of Financial Studies* 24 (12): 4037–90.
Chernenko, Sergey, Josh Lerner, and Yao Zeng. 2017. "Mutual Funds as Venture Capitalists? Evidence from Unicorns." NBER Working Paper 23981.
Cornelli, Francesca, and Oved Yosha. 2003. "Stage Financing and the Role of Convertible Securities." *Review of Economic Studies* 70 (1): 1–32.
Da Rin, Marco, Thomas Hellmann, and Manju Puri. 2013. "A Survey of Venture Capital Research." In *Handbook of the Economics of Finance*, Volume 2, Part A, edited by George Constantinides, Milton Harris, and René Stulz, 573–648. Amsterdam: Elsevier.
de Roover, Raymond. 1963. "The Organization of Trade." In *The Cambridge Economic History of Europe: Volume III: Economic Organization and Policies in the Middle Ages*, edited by M.M. Postan, E.E. Rich, and Edward Miller, chapter 2. Cambridge: Cambridge University Press.
Eldar, Ofar, Yael Hochberg, Lubomir Litov. 2020. "The Rise of Dual-Class Stock IPOs and Venture Capital Financing." Unpublished.
Ewens, Michael, and Matt Marx. 2018. "Founder Replacement and Startup Performance." *Review of Financial Studies* 31 (4): 1532–65.
Ewens, Michael, and Joan Farre-Mensa. Forthcoming. "The Deregulation of the Private Equity Markets and the Decline in IPOs." *Review of Financial Studies*.
Ewens, Michael, Ramana Nanda, and Matthew Rhodes-Kropf. 2018. "Cost of Experimentation and the Evolution of Venture Capital." *Journal of Financial Economics* 128 (3): 422–42.
Ewens, Michael, and Matthew Rhodes-Kropf. 2015. "Is a VC Partnership Greater than the Sum of Its Partners?" *Journal of Finance* 70 (3): 1081–1113.
Ewens, Michael, and Richard Townsend. 2020. "Are Early Stage Investors Biased Against Women?" *Journal of Financial Economics* 135 (3): 653–77.
Fang, Lily, Victoria Ivashina, and Josh Lerner. 2015. "The Disintermediation of Financial Markets: Direct Investing in Private Equity." *Journal of Financial Economics* 116 (1): 160–78.
Farre-Mensa, Joan, Deepak Hegde, and Alexander Ljungqvist. 2020. "What Is a Patent Worth? Evidence from the U.S. Patent 'Lottery.'" *Journal of Finance* 75 (2): 639–82.
Florida, Richard, and Ian Hathaway. 2018. *The Rise of the Startup City*. Washington: Center for American Entrepreneurship. AQ: Please provide link
Gao, Xiaohui, Jay R. Ritter, and Zhongyan Zhu. 2013. "Where Have All the IPOs Gone?" *Journal of Financial and Quantitative Analysis* 48 (6): 1663–92.
Glaeser, Edward, and Naomi Hausman. 2020. "The Spatial Mismatch between Innovation and Joblessness." *Innovation Policy and the Economy* 20: 233–99.
Gompers, Paul A. 1995. "Optimal Investment, Monitoring, and the Staging of Venture Capital." *Journal of Finance* 50 (5): 1461–90.
Gompers, Paul A., Will Gornall, Steven N. Kaplan, and Ilya A. Strebulaev. 2020. "How Do Venture Capitalists Make Decisions?" *Journal of Financial Economics* 135 (1): 169–90.
Gompers, Paul A., and Sophie Q. Wang. 2017. "And the Children Shall Lead: Gender Diversity and Performance in Venture Capital." NBER Working Paper 23454.
Gornall, Will, and Ilya A. Strebulaev. 2015. "The Economic Impact of Venture Capital: Evidence from Public Companies." Stanford University Graduate School of Business Research Paper 15–55.
Gompers, Paul, and Josh Lerner. 2001a. *The Money of Invention: How Venture Capital Creates New Wealth*. Boston: Harvard Business School Press.
Gompers, Paul, and Josh Lerner. 2001b. "The Venture Capital Revolution." *Journal of Economic Perspectives* 15 (2): 145–68.
Gonzalez-Uribe, Juanita, and Michael Leatherbee. 2017. "The Effects of Business Accelerators on Venture Performance: Evidence from Start-up Chile." *The Review of Financial Studies* 31 (4): 1566–1603.
Hall, Robert E., and Susan E. Woodward. 2004. "Benchmarking the Returns to Venture." NBER 10202.
Hall, Robert E., and Susan E. Woodward. 2010. "The Burden of the Nondiversifiable Risk of Entrepreneurship." *American Economic Review* 100 (3): 1163–94.
Harris, Robert S., Tim Jenkinson, Steven N. Kaplan, and Ruediger Stucke. 2014. "Has Persistence Persisted in Private Equity? Evidence from Buyout and Venture Capital Funds." Fama-Miller Working Paper 2304808.
Hellmann, Thomas. 1998. "The Allocation of Control Rights in Venture Capital Contracts." *Rand Journal*

of Economics 29 (1): 57–76.
Hellmann, Thomas, and Manju Puri. 2000. "The Interaction between Product Market and Financing Strategy: The Role of Venture Capital." *Review of Financial Studies* 13 (4): 959–84.
Hellmann, Thomas, and Manju Puri. 2002. "Venture Capital and the Professionalization of Start-up Firms: Empirical Evidence." *Journal of Finance* 57 (1): 169–97.
Hochberg, Yael. 2016. "Accelerating Entrepreneurs and Ecosystems: The Seed Accelerator Model." *Innovation Policy and the Economy* 16: 25–51.
Hochberg, Yael V., Alexander Ljungqvist, and Yang Lu. 2007. "Whom You Know Matters: Venture Capital Networks and Investment Performance." *Journal of Finance* 62 (1): 251–301.
Howell, Sabrina T. 2017. "Financing Innovation: Evidence from R&D Grants." *American Economic Review* 107 (4): 1136–64.
Howell, Sabrina T., Josh Lerner, Ramana Nanda, and Richard Townsend. 2020. "Financial Distancing: How Venture Capital Follows the Economy Down and Curtails Innovation." NBER Working Paper 27150.
Howell, Sabrina T., Marina Niessner, and David Yermack. Forthcoming. "Initial Coin Offerings: Financing Growth with Cryptocurrency Token Sales." *Review of Financial Studies*.
Hsu, David. 2004. "What Do Entrepreneurs Pay for Venture Capital Affiliation?" *The Journal of Finance* 59 (4): 1805–44.
Iliev, Peter. 2010. "The Effect of SOX Section 404: Costs, Earnings Quality, and Stock Prices." *Journal of Finance* 65 (3): 1163–96.
Ivashina, Victoria, and Josh Lerner. 2019. *Patent Capital: The Challenges and Promises of Long-Term Investing*. Princeton: Princeton University Press.
Iyer, Rajkamal, Asim Ijaz Khwaja, Erzo F. P. Luttmer, and Kelly Shue. 2016. "Screening Peers Softly: Inferring the Quality of Small Borrowers." *Management Science* 62 (6): 1554–77.
Janeway, William H. 2018. *Doing Capitalism in the Innovation Economy: Markets, Speculation and the State*. Cambridge: Cambridge University Press.
Kaplan, Steven N., and Antoinette Schoar. 2005. "Private Equity Performance: Returns, Persistence, and Capital Flows." *Journal of Finance* 60 (4): 1791–1823.
Kaplan, Steven N., Berk Sensoy, and Per Strömberg. 2009. "Should Investors Bet on the Jockey or the Horse: Evidence from the Evolution of Firms from Early Business Plans to Public Companies." *Journal of Finance* 64 (1): 75–115.
Kaplan, Steven N., and Per Strömberg. 2003. "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts." *Review of Economic Studies* 70 (2): 281–315.
Kaplan, Steven N., and Per Strömberg. 2004. "Characteristics, Contracts, and Actions: Evidence from Venture Capitalist Analyses." *Journal of Finance* 59 (5): 2177–2210.
Kerr, William R., Ramana Nanda, and Matthew Rhodes-Kropf. 2014. "Entrepreneurship as Experimentation." *Journal of Economic Perspectives* 28 (3): 25–48.
Kerr, William R., Josh Lerner, and Antoinette Schoar. 2014. "The Consequences of Entrepreneurial Finance: A Regression Discontinuity Analysis." *Review of Financial Studies* 27 (1): 20–55.
Korteweg, Arthur, and Morten Sørensen. 2017. "Skill and Luck in Private Equity Performance." *Journal of Financial Economics* 124 (3): 535–62.
Kortum, Samuel, and Josh Lerner. 2000. "Assessing the Impact of Venture Capital on Innovation." *Rand Journal of Economics* 31 (4): 674–92.
Lerner, Josh. 1995. "Venture Capitalists and the Oversight of Private Firms." *Journal of Finance* 50 (1): 301–18.
Lerner, Josh. 1999. "The Government as Venture Capitalist: The Long-Run Effects of the SBIR Program." *Journal of Business* 72: 285–318.
Lerner, Josh. 2012. *The Architecture of Innovation: The Economics of Creative Organizations*. Boston: Harvard Business Review Press.
Lerner, Josh, Jason Mao, Antoinette Schoar, and Nan R. Zhang. 2018. "Investing Outside the Box: Evidence from Alternative Vehicles in Private Capital." NBER Working Paper 24941.
Lerner, Josh, and Julie Wulf. 2007. "Innovation and Incentives: Evidence from Corporate R&D." *Review of Economics and Statistics* 89 (4): 634–44.
Lin, Lin. 2017. "Managing the Risks of Equity Crowdfunding: Lessons from China." *Journal of Corporate Law Studies* 17 (2): 327–66.
Lopez Robert S., and Irving W. Raymond. 1955. *Medieval Trade in the Mediterranean World: Illustrative Documents Translated with Introductions and Notes*. New York: Columbia University Press.
Ma, Song. 2020. "The Life Cycle of Corporate Venture Capital." *Review of Financial Studies* 33 (1): 358–94.
Metrick, Andrew, and Ayako Yasuda. 2010. "The Economics of Private Equity Funds." *Review of Financial*

Studies 23 (6): 2303–41.

Mollick, Ethan. 2014. "The Dynamics of Crowdfunding: An Exploratory Study." *Journal of Business Venturing* 29 (1): 1–16.

Murray, Sarah. 2020. "Philanthropists Play a Crucial Role in Developing Vaccines." *Financial Times*, May 21, https://www.ft.com/content/847a9052-6847-11ea-a6ac-9122541af204?shareType=nongift.

Nanda, Ramana, and Matthew Rhodes-Kropf. 2017. "Financing Risk and Innovation." *Management Science* 63 (4): 901–18.

Nanda, Ramana, Sampsa Samila, and Olav Sorenson. 2020. "The Persistent Effect of Initial Success: Evidence from Venture Capital." *Journal of Financial Economics* 137 (1): 231–48.

National Venture Capital Association. 2020. "NVCA 2020 Yearbook." *NVCA*. https://nvca.org/wp-content/uploads/2020/04/NVCA-2020-Yearbook.pdf. (accessed April 1, 2020).

Neher, Darwin V. 1999. "Staged Financing: An Agency Perspective." *Review of Economic Studies* 66 (2): 255–74.

Nicholas, Tom. 2019. *VC: An American History* Cambridge: Harvard University Press.

Puri, Manju, and Rebecca Zarutskie. 2012. "On the Lifecycle Dynamics of Venture-Capital- and Non-Venture-Capital-Financed Firms." *Journal of Finance* 67 (6): 2247–93.

Ries, Eric. 2011. *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. New York: Crown Publishing Group.

Securities Industry and Financial Markets Association (SIFMA). 2019. *Capital Markets Fact Book*. New York: SIFMA.

Sørensen, Morten. 2007. "How Smart Is the Smart Money? A Two-Sided Matching Model of Venture Capital." *Journal of Finance* 62 (6): 2725–62.

US Small Business Administration. 2018. "SBIR/STTR Annual Report 2017." Washington, D.C.: U.S. Small Business Administration.

Zetzsche, Dirk Andreas, Ross Buckley, Douglas Arner, and Linus Föhr. 2018. "The ICO Gold Rush: It's a Scam, It's a Bubble, It's a Super Challenge for Regulators." European Banking Institute Working Paper Series 18/2018.

Journal of Economic Perspectives—Volume 34, Number 3—Summer 2020—Pages 262–269

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by e-mail at taylort@macalester.edu, or c/o *Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., St. Paul, MN 55105.

Smorgasbord

In the Presidential Address for the Eastern Economic Association, Edward L. Glaeser considers "Urbanization and Its Discontents" (*Eastern Economic Journal*, April 2020, 46:191–218, https://link.springer.com/article/10.1057/s41302-020-00167-3). "The industrial jobs that had once been the backbone of urban economies did not return. Instead, human capital-intensive business services became the new export industries for urban areas. Financial services expanded enormously in urban America from 1980 to 2007. At its height in 2007, finance and insurance generated over forty percent of the total payroll on the island of Manhattan. The urban edge in transferring knowledge is particularly valuable in finance because, a bit of extra information can make millions for a trader in minutes. . . . Why didn't

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota. He blogs at http://conversableeconomist.blogspot.com.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.34.3.262.

improvements in electronic communication make face-to-face contact obsolete? While e-mail is possible almost everywhere, face-to-face interactions generate a richer information flow that includes body language, intonation and facial expression. As the world became more complex, the value of intense communication also increases. Physical immersion in an informationally intense environment, such as trading floor or an academic seminar, generates a rush of information that is hard to duplicate online. Moreover, dense environments facilitate random personal interactions that can create serendipitous flows of knowledge and collaborative creativity. The knowledge-intensive nature of the urban resurgence helps to explain why educated cities have done much better than uneducated cities. . . . Why has urban success been accompanied by so much discontent? The most natural explanation is that the success of private enterprise in cities has not been accompanied by sufficient development of public capacity. The public sector has often focused on limiting urban change, rather than working to improve the urban experience. In many cases, this focus reflects the political priorities of empowered insiders." Glaeser's address complements the papers on urban economics in this issue.

Andrea M. Headley and James E. Wright, II, look at "National Police Reform Commissions: Evidence-Based Practices or Unfulfilled Promises?" (*Review of Black Political Economy*, December 2019, 46:4, pp. 277–305, https://journals.sagepub.com/doi/full/10.1177/0034644619873074). "COP [community-oriented policing] is a promising practice to build police–community relations, particularly for communities of color. . . . The research has generally shown COP positively affects community perceptions and attitudes and thus builds relations, whereas such strategies have very limited, if any, effects on reducing crime. That being said, there is no clear guidance as to which specific features of COP make the most difference and/or how best to implement COP strategies." "[M]ost of the national commissions have recommended a commitment to improving the quality of the workforce by hiring more people of color and women, implementing educational requirements for officers, increasing hiring standards, and providing more effective training. . . . By and large, the use of enhanced or more stringent hiring standards and prescreening assessments to improve professionalism and the quality of the police force cannot be supported by the evidence herein. Police departments across the country are realizing the need to expand their hiring pool while also acknowledging some of the harms that have been done to keep people of color and women out of policing (whether intentionally or not) . . . Training has been one of the most commonly used ways to respond to crises in the policing profession in hopes to affect police behavior. Unfortunately, with the lack of consistency in training across police departments, scholarship has not rigorously or systematically been able to examine the impacts of various types of trainings." "Police culture may influence and reinforce certain types of behavior and/or beliefs in officers that are counter to the police reforms existing at the structural or administrative level."

Kass Forman, Sean Dougherty, and Hansjörg Blöchliger provide an overview in "Synthesising Good Practices in Fiscal Federalism: Key recommendations

from 15 years of country surveys" (OECD Economic Policy Paper #28, April 2020, https://www.oecd.org/china/synthesising-good-practices-in-fiscal-federalism-89cd0319-en.htm). "Fiscal federalism refers to the distribution of taxation and spending powers across levels of government. Through decentralisation, governments can bring public services closer to households and firms, allowing better adaptation to local preferences. However, decentralisation can also make intergovernmental fiscal frameworks more complex and risk reinforcing interregional inequality unless properly designed. Accordingly, several important trade-offs emerge from the devolution of tax and spending powers. . . . OECD research has found a broadly positive relationship between revenue decentralisation and growth, with spending decentralisation demonstrating a weaker effect . . . [D]ecentralisation appears to reduce the gap between high and middle-income households but may leave low incomes behind, especially where jurisdictions have large tax autonomy . . . In healthcare, research suggests costs fall and life expectancy rises with moderate decentralisation, but the opposite effects hold once decentralisation becomes excessive . . . With respect to educational attainment, . . . a 10 percentage point increase in the sub-national revenue share improves PISA scores by 6 percentage points . . . "

Stephen Broadberry and Mark Harrison have edited an ebook with 16 short chapters: *The Economics of the Second World War: Seventy-Five Years On* (May 2020, CEPR Press, https://voxeu.org/content/new-ebook-economics-second-world-war-seventy-five-years, free registration required). As one example, Phillips Payson O'Brien contributes "How the War Was Won." "Looking at the war this way allows us to reframe our understanding of what a battle was in the Second World War. Instead of battles being fixed on well-known pieces of earth, air-sea weaponry was constantly in action in battlefields thousands of miles long and many miles in depth—what should be called the Air-Sea Super Battlefield. Victory in this super-battlefield led to victory in the war. . . . Instead of waiting to destroy Axis equipment on the traditional battlefield, Allied air-sea weaponry destroyed it en masse before it could ever be used in action, determining the result of every 'battle' long before it was fought. . . . First, there is pre-production destruction, which prevented weapons from being built. This was done most efficiently to both Germany and Japan by depriving them of the ability to move raw materials. . . . The second phase is direct production destruction—destroying the facilities to make weapons in Germany and Japan. . . . The truth was that these attacks were not as effective as hoped for, as strategic bombing was not accurate enough to completely wipe out facilities (until 1944). That being said, the losses from bombing were greater than those arising in land battles. . . . Finally, there were deployment losses. Getting weapons from the factory to the front was no easy feat. It normally required movement over hundreds or thousands of miles using shipping or rail lines that were vulnerable to attack. Aircraft had to be flown, often by inexperienced pilots, over the open ocean in or through difficult weather conditions. By 1943, . . . the Axis were losing as many aircraft deploying to the front as in direct combat. At times, Japan's losses outside combat were up to twice those lost fighting . . . This was the true battlefield of the

Second World War, a massive air-sea super battlefield that stretched for thousands of miles not only of traditional front but of depth and height."

Jonathan Vespa, Lauren Medina, and David M. Armstrong have written "Demographic Turning Points for the United States: Population Projections for 2020 to 2060" (US Census Bureau, Report P25-1144, revised February 2020, https://www.census.gov/library/publications/2020/demo/p25-1144.html). "The year 2030 marks a demographic turning point for the United States. Beginning that year, all baby boomers will be older than 65. This will expand the size of the older population so that one in every five Americans is projected to be retirement age. Later that decade, by 2034, we project that older adults will outnumber children for the first time in U.S. history. The year 2030 marks another demographic first for the United States. That year, because of population aging, immigration is projected to overtake natural increase (the excess of births over deaths) as the primary driver of population growth for the country. As the population ages, the number of deaths is projected to rise substantially, which will slow the country's natural growth. As a result, net international migration is projected to overtake natural increase, even as levels of migration are projected to remain relatively flat. These three demographic milestones are expected to make the 2030s a transformative decade for the U.S. population." For a global perspective on aging and demographic change, the March 2020 issue of *Finance and Development* has a symposium of eight short articles at https://www.imf.org/external/pubs/ft/fandd/2020/03/index.htm.

Lucian Bebchuk and Scott Hirst address "Index Funds and the Future of Corporate Governance: Theory, Evidence, and Policy" (*Columbia Law Review*, December 2019, pp. 2029–2145, https://columbialawreview.org/content/index-funds-and-the-future-of-corporate-governance-theory-evidence-and-policy/). "We put forward a set of reforms that policymakers should consider in order to address the incentives of index fund managers to underinvest in stewardship, their incentives to be excessively deferential to corporate managers, and the continuing rise of index investing. . . . These problems are expected to remain a significant aspect of the corporate governance landscape and should be the subject of close attention by policymakers, market participants, and scholars." The essay can be read as a follow-up to their essay "The Specter of the Giant Three" (*Boston University Law Review*, May 2019, 99:3, pp. 721–42, https://www.bu.edu/bulawreview/files/2019/06/BEBCHUK-HIRST-1.pdf), or as a follow-up to the article by Bebchuk and Hirsi (with Alma Cohen) in the Summer 2017 issue of this journal.

There are waves of new economic research in response to the COVID-19 pandemic. Rather than try to mention a couple of working papers here, which may be superceded by the time this issue is published, I'll just point out that the National Bureau of Economic Research has made its working papers related to pandemic freely available at https://www.nber.org. Also, the Centre for Economic Policy Research launched an online *COVID Economics* journal in late March, which has already published more than 30 issues that typically include 6–8 papers each at https://cepr.org/content/covid-economics/.

Interviews with Economists

Douglas Clement at the Minneapolis Federal Reserve offers one of his characteristically excellent interviews, this one with Emi Nakamura, titled "On price dynamics, monetary policy, and this 'scary moment in history'" (Federal Reserve Bank of Minneapolis. May 6, 2020, https://www.minneapolisfed.org/article/2020/emi-nakamura-interview-on-price-dynamics-monetary-policy-and-this-scary-moment-in-history). "You might think that it's very easy to go out there and figure out how much rigidity there is in prices. But the reality was that at least until 20 years ago, it was pretty hard to get broad-based price data. In principle, you could go into any store and see what the prices were, but the data just weren't available to researchers tabulated in a systematic way. . . . Once macroeconomists started looking at data for this broad cross section of goods, it was obvious that pricing behavior was a lot more complicated in the real world than had been assumed. If you look at, say, soft drink prices, they change all the time. But the question macroeconomists want to answer is more nuanced. We know that Coke and Pepsi go on sale a lot. But is that really a response to macroeconomic phenomena, or is that something that is, in some sense, on autopilot or preprogrammed? Another question is: When you see a price change, is it a response, in some sense, to macroeconomic conditions? We found that, often, the price is simply going back to exactly the same price as before the sale. That suggests that the responsiveness to macroeconomic conditions associated with these sales was fairly limited. . . . One of the things that's been very striking to me in the recent period of the COVID-19 crisis is that even with incredible runs on grocery products, when I order my online groceries, there are still things on sale. Even with a shock as big as the COVID shock, my guess is that these things take time to adjust. . . . The COVID-19 crisis can be viewed as a prime example of the kind of negative productivity shock that neoclassical economists have traditionally focused on. But an economy with price rigidity responds much less efficiently to that kind of an adverse shock than if prices and wages were continuously adjusting in an optimal way."

Isaac Chotiner has a short interview with Paul Romer ("Paul Romer's Case for Nationwide Coronavirus Testing," *New Yorker*, May 3, 2020, https://www.newyorker.com/news/q-and-a/paul-romer-on-how-to-survive-the-chaos-of-the-coronavirus). "The gains from specialization go all the way back to Adam Smith. He talked about the advantage of a bigger market being that we could have a finer division of labor and be more specialized. There's this great story about the pin factory where people do various different pieces of the job of making pins. So, we've been very attuned to the efficiency gains that come from finer and finer division of labor and specialization. What we've underestimated is the systemic risk that that very finely tuned system of specialization exposes us to. And so I think we will start to ask whether there are ways that we could build some more robustness into our whole system. If I can use an analogy, Netflix used this thing they called the Chaos Monkey, which would go in and just break servers, break routers, just take them offline and then make sure that the Netflix infrastructure system could still keep working. I think,

from a public-policy perspective, it'd be good if we started having some drills where we just break things, like, 'O.K., you can't import that input into your pharmaceutical process for six months,' or, 'You can't rely on this mechanism.' We may need a little bit of a Chaos Monkey to help make sure that we're all building a little bit more resiliency into the things that we do."

Irwin Stelzer and Jeffrey Gedmin interview Lawrence Summers ("How to Fix Globalization—for Detroit, Not Davos" *The American Interest*, May 22, 2020, https://www.the-american-interest.com/2020/05/22/how-to-fix-globalization-for-detroit-not-davos/). On globalization: "Someone put it to me this way: First, we said that you are going to lose your job, but it was okay because when you got your new one, you were going to have higher wages thanks to lower prices because of international trade. Then we said that your company was going to move your job overseas, but it was really necessary because if we didn't do that, then your company was going to be less competitive. Now we're saying that we have to cut the taxes on those companies and cut the calculus class from your kid's high school, because otherwise we won't be able to attract companies to the United States, and you have to pay higher taxes and live with fewer services. At a certain point, people say, 'This whole global thing doesn't work for me,' and they have a point." On government debt: "The deepest truth about debt is that you can't evaluate borrowing without knowing what it's going to be used for. Borrowing to invest in ways that earn a higher return than the cost of borrowing, and provide the wherewithal for debt service with an excess left-over, is generally a good and sustainable thing. Borrowing to finance consumption, leaving no return to cover debt service, is generally an unsustainable and problematic thing. . . . I think we need to be very careful, with respect to the expectation that we now seem to be setting of having government cover all the losses associated with the COVID period. . . . Looking towards an economy that is going to be very different than the one we had before COVID, we cannot aspire to maintain every job or every enterprise with a compensation program indefinitely. So as I look at the 30 percent of GDP deficit that we are running in Quarters Three and Four of Fiscal 2020, I don't think that can be sustained over a multi-year period."

Merle van den Akker has an "Interview with Colin Camerer" ("Money on the Mind," April 6, 2020, https://www.moneyonthemind.org/post/interview-with-colin-camerer). Here is some advice from Camerer for an aspiring behavioral economist: "First, you need to know the 'rules' of economics—the basic canon and methods—very well. . . . To break the rules you need to know the rules. Second, in my opinion, if you want to succeed in behavioral economics it is a big help to be very fluent in an adjacent social science. A lot of behavioral economics is in the business of importing ideas and translating them, redesigning and "selling" them inside economics. So you need to become bilingual and know what psychology, or neuroscience, media studies, or whatever, is solid, and has a long good empirical pedigree. Figuring that out can be difficult. Third, nowadays you really should be able to do lab (and online) experiments, know about quasi-experimental designs (IV, diff-in-diff, regression discontinuity) and know some machine learning. It is often said that most of the methods you will use in your long research career are

those you learned in graduate school. It is like packing for a long, long trip to a place where there are no stores in case you forgot to pack anything. Fill that backpack with methods."

David A. Price acts as the interlocutor in "Interview: Joshua Angrist" (*Econ Focus*: Federal Reserve Bank of Richmond, First Quarter 2020, pp. 18–22, https://www.richmondfed.org/publications/research/econ_focus/2020/q1/interview). "[O]ne of my favorite examples for teaching regression is a paper by Alan Krueger and Stacy Dale that looks at the effects of going to a more selective college. It turns out that if you got into MIT or Harvard, it actually doesn't matter where you go. Alan and Stacy showed that in two very clever, well-controlled studies. And Jack Mountjoy, in a paper with Brent Hickman, just replicated that for a much larger sample. There isn't any earnings advantage from going to a more selective school once you control for the selection bias. So there's also an elite illusion at the college level, which I think is more important to upper-income families, because they're desperate for their kids to go to the top schools. So desperate, in fact, that a few commit criminal fraud to get their kids into more selective schools."

Discussion Starters

Noel-Ann Bradshaw discusses some work by the first female member of Britain's Royal Statistical Society in "Florence Nightingale (1820–1910): An Unexpected Master of Data" (*Patterns*, May 2020, https://www.cell.com/patterns/fulltext/S2666-3899(20)30041-6). "[Nightingale] became fascinated that the mortality rate among soldiers stationed at home was higher than the mortality rate of ordinary British men, despite soldiers being healthier at the start of their careers. She used data to examine the cause, concluding that the problem was poor sanitation and over-crowding of military barracks, encampments, and hospitals that exacerbated the spread of disease. She drew many graphs depicting this, including Figure 1, which shows five circles filled with hexagons representing the space between people. The first three circles show how closely packed the army would be in the Quartermaster General's camp plans, while the last two circles show how densely packed the inner city of London currently was and the population of London in general. This comparison made it obvious to anyone that the Quartermaster General's proposition for encampment was going to be problematic given how unhealthy densely populated areas of London were. . . . She went on to forecast the efficiency of the army if the soldiers were as healthy as the rest of the men in the UK. This graph was way ahead of its time (Figure 2). On the left she displayed the current situation, showing the effectiveness of the British Army in terms of the numbers who were ill, invalided, etc. On the right she graphed the potential effectiveness of the army if the soldiers were as healthy as the general male population. By forecasting this potential effectiveness, she emphasized how the army at rest were experiencing higher degrees of mortality that the general male population."

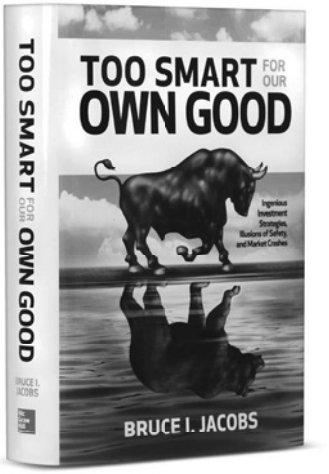
Randal O'Toole makes his case for "Transit: The Urban Parasite" (Cato Institute, Policy Analysis #889, April 20, 2020, https://www.cato.org/publications/policy-analysis/transit-urban-parasite). "Data released by the Federal Transit Administration in December 2019 indicate that 2018 transit ridership fell in 40 of the nation's top 50 urban areas, and, over the past five years ridership has fallen in 44 of those 50 urban areas. . . . These declines have taken place in spite of huge increases in spending on public transit. In 2018 alone, subsidies to transit grew by 7.4 percent, increasing from $50.5 billion to $54.3 billion. . . . [T]he justifications for spending this much money subsidizing a declining industry are disappearing. Most low-income workers have given up on transit as a method of commuting and have purchased cars. . . . In all but a handful of urban areas, transit uses more energy and emits more greenhouse gases per passenger mile than the average automobile. Far from relieving congestion, transit agencies are seeking to increase congestion in order to promote their businesses. . . . Transit advocates have reached the point where they act as though the purpose of cities and their residents is to benefit transit. In fact, transit should benefit residents by enhancing their mobility and well-being. If transit is not doing that, and people no longer value it, then it should not be subsidized."

Dan Lovallo, Tim Koller, Robert Uhlaner, and Daniel Kahneman argue "Your Company Is Too Risk-Averse" (*Harvard Business Review*, March-April 2020, https://hbr.org/2020/03/your-company-is-too-risk-averse). "In current practice, however, executives in large corporations are reluctant to propose and advocate for risky projects. They quash new ideas in favor of marginal improvements, cost-cutting, and "safe" investments. Research studies long ago established this pattern. In a classic HBR article, for example, Syracuse University professor Ralph O. Swalm presented the results of a remarkable study of risk attitudes among 100 executives. He concluded that the findings "do not portray the risk-takers we hear so much of in industrial folklore. They portray decision-makers quite unwilling to take what, for the company, would seem to be rather attractive risks." Our research confirms that this pattern persists. . . . CEOs are evaluated on their long-term performance, but managers at lower levels essentially bet their careers on every decision they make—even if outcomes are negligible to the corporation as a whole."

Too Smart for Our Own Good

Ingenious Investment Strategies, Illusions of Safety, and Market Crashes

Bruce I. Jacobs



One of today's leading financial thinkers, Bruce I. Jacobs, examines recent financial crises—including the 1987 stock market crash, the 1998 collapse of the hedge fund Long-Term Capital Management, the 2007–2008 credit crisis, and the European debt crisis—and reveals the common threads that explain these market disruptions. In each case, investors in search of safety were drawn to novel strategies that were intended to reduce risk but actually magnified it—and blew up markets. Until we manage risk in responsible ways, major crises will always be just around the bend. *Too Smart for Our Own Good* is a big step toward smarter investing—and a better financial future for everyone.

"This is a highly readable account of a series of innovations that proved too clever by half."
—Richard J. Herring, Professor of Finance, and Director, Wharton Financial Institutions Center, The Wharton School

"Bruce Jacobs has produced an important and timely book that explains the common themes that underlie these disruptive events and offers the possibility of avoiding them in the future. It will be of inestimable, and equal, value to practitioners, regulators, and the academic community."
—Richard Lindsey, former Director of Market Regulation and Chief Economist, SEC

"This is the book investors should read today to be prepared for the next crash, which is certain to come."
—Edward M. Miller, Professor of Economics and Finance, University of New Orleans

"Bruce Jacobs explains in clear and often gripping ways how leverage, opacity, and complex investment strategies contributed to market meltdowns. Anyone who wants markets to be safer and more stable should harken to Jacobs's words of wisdom."
—Frank Partnoy, Author of *F.I.A.S.C.O.* and *Infectious Greed*, and Professor of Law, University of California, Berkeley

"Bruce Jacobs does a splendid job of connecting the dots of the causes of crises and suggests how we can think about the daunting task of 'taming the tempest.'"
—Hersh Shefrin, Professor of Finance, Leavey School of Business, Santa Clara University

"In this very thoughtful and comprehensive book, Bruce Jacobs takes the reader on a tour of the financial markets and the market crises we have lived through. I highly recommend this well researched and written book."
—William T. Ziemba, Professor Emeritus, University of British Columbia

About the Author

Bruce I. Jacobs is co-founder, co-chief investment officer, and co-director of research at Jacobs Levy Equity Management. He is co-author, with Ken Levy, of *Equity Management: The Art and Science of Modern Quantitative Investing*. Jacobs serves on the Advisory Boards of the *Journal of Portfolio Management* and *Journal of Financial Data Science*, and has served on the *Financial Analysts Journal* Advisory Council. He holds a Ph.D. in finance from The Wharton School.

Mc Graw Hill Education
Visit: www.mhprofessional.com
ISBN: 9781260440546
Available in print, ebook, and audiobook formats

From the American Economic Association

RESEARCH HIGHLIGHTS

A Convenient Way to Monitor Key Economics Research and Emerging Topics Being Published in AEA Journals

- *Article Summaries on Key Topics*
- *Dedicated Web Content Editor*
- *Weekly Updates*
- *Interactive Charts and Graphs*
- *Links to Related Materials*

Podcast Now Available!

AMERICAN ECONOMIC ASSOCIATION

Find the latest complimentary Research Highlights at

www.aeaweb.org/research

www.aeaweb.org/research/podcasts

@aeajournals

Committee on the Status of Minority Groups in the Economics Profession

AMERICAN ECONOMIC ASSOCIATION

• Established 1968 •

SUPPORTING DIVERSITY IN ECONOMICS

The Committee on the Status of Minority Groups in the Economics Profession (CSMGEP) was established by the American Economic Association (AEA) in 1968 to increase the representation of minorities in the economics profession, primarily by broadening opportunities for the training of underrepresented minorities.

CSMGEP Programs

- Summer Economics Fellows Program
- Mentoring Program
- Summer Training Program

www.csmgep.org

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary access to JEP articles, go to the AEA website: http://www.aeaweb.org. Annual dues for regular membership are $24.00, $34.00, or $44.00, depending on income; for an additional fee, you can receive this journal, or any of the Association's journals, in print. Change of address notice must be received at least six weeks prior to the publication month.

Copyright © 2020 by the American Economic Association. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation, including the name of the author. Copyrights for components of this work owned by others than AEA must be honored. Abstracting with credit is permitted. The author has the right to republish, post on servers, redistribute to lists, and use any component of this work in other works. For others to do so requires prior specific permission and/or a fee. Permissions may be requested from the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203; email: aeainfo@vanderbilt.edu.

MIX
Paper from responsible sources
FSC™ C132124
FSC
www.fsc.org

Founded in 1885

EXECUTIVE COMMITTEE

Elected Officers and Members

President
JANET L. YELLEN, The Brookings Institution
President-elect
DAVID CARD, University of California, Berkeley
Vice Presidents
JANICE EBERLY, Northwestern University
OLIVIA S. MITCHELL, University of Pennsylvania
Members
ADRIANA LLERAS-MUNEY, University of California, Los Angeles
BETSEY STEVENSON, University of Michigan
MARTHA BAILEY, University of Michigan
SUSANTO BASU, Boston College
LISA D. COOK, Michigan State University
MELISSA S. KEARNEY, University of Maryland
Ex Officio Members
OLIVIER BLANCHARD, Peterson Institute for International Economics
BEN S. BERNANKE, The Brookings Institution

Appointed Members

Editor, The American Economic Review
ESTHER DUFLO, Massachusetts Institute of Technology
Editor, The American Economic Review: Insights
AMY FINKELSTEIN, Massachusetts Institute of Technology
Editor, The Journal of Economic Literature
STEVEN N. DURLAUF, University of Chicago
Editor, The Journal of Economic Perspectives
ENRICO MORETTI, University of California, Berkeley
Editor, American Economic Journal: Applied Economics
BENJAMIN OLKEN, Massachusetts Institute of Technology
Editor, American Economic Journal: Economic Policy
ERZO F.P. LUTTMER, Dartmouth College
Editor, American Economic Journal: Macroeconomics
SIMON GILCHRIST, New York University
Editor, American Economic Journal: Microeconomics
LEEAT YARIV, Princeton University
Secretary-Treasurer
PETER L. ROUSSEAU, Vanderbilt University

OTHER OFFICERS

Director of AEA Publication Services
ELIZABETH R. BRAUNSTEIN
Counsel
LAUREN M. GAFFNEY, Bass, Berry & Sims PLC
Nashville, TN

ADMINISTRATORS

Director of Finance and Administration
BARBARA H. FISER
Convention Manager
GWYN LOFTIS

The Journal of

Economic Perspectives

Summer 2020, Volume 34, Number 3

Symposia

Productivity Advantages of Cities

Place-Based Policies

Cities in Developing Countries

Articles

AMERICAN ECONOMIC ASSOCIATION