

*The Journal of*

*Economic Perspectives*

*A journal of the  
American Economic Association*

*Winter 2024*

# The Journal of Economic Perspectives

*A journal of the American Economic Association*

## **Editor**

Heidi Williams, Dartmouth College

## **Coeditors**

Jonathan Parker, Massachusetts Institute of Technology

Nina Pavcnik, Dartmouth College

## **Associate Editors**

Panle Jia Barwick, University of Wisconsin-Madison

Anusha Chari, University of North Carolina

Karen Clay, Carnegie Mellon University

David Deming, Harvard University

Andrea Eisfeldt, University of California at Los Angeles

Rachel Glennerster, University of Chicago

Louis Kaplow, Harvard University

Imran Rasul, University College London

Juan Carlos Serrato, Duke University

Lise Vesterlund, University of Pittsburgh

Jeffrey Wooldridge, Michigan State University

## **Data Editor**

Lars Vilhuber

## **Managing Editor**

Timothy Taylor

## **Assistant Managing Editor**

Bradley Waldruff

---

### *Editorial offices:*

Journal of Economic Perspectives  
*American Economic Association Publications*  
2403 Sidney St., #260  
Pittsburgh, PA 15203  
*email: [jep@aea-pubs.org](mailto:jep@aea-pubs.org)*

---

The *Journal of Economic Perspectives* gratefully acknowledges the support of Macalester College. Registered in the US Patent and Trademark Office (®).

Copyright © 2024 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), Winter 2024, Vol. 38, No. 1. The JEP is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. 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 USA.

*The Journal of*  
***Economic Perspectives***

---

**Contents**

*Volume 38 • Number 1 • Winter 2024*

---

**Symposia**

***The Global Market for Talent***

Britta Glennon, “Skilled Immigrants, Firms, and the Global Geography of Innovation” . . . . . 3

Francesco Lissoni and Ernest Miguelez, “Migration and Innovation: Learning from Patent and Inventor Data” . . . . . 27

***Taxation and Developing Countries***

Pierre Bachas ⊕ Anders Jensen ⊕ Lucie Gadenne, “Tax Equity in Low- and Middle-Income Countries” . . . . . 55

Oyebola Okunogbe and Gabriel Tourek, “How Can Lower-Income Countries Collect More Taxes? The Role of Technology, Tax Agents, and Politics” . . . . . 81

Anne Brockmeyer ⊕ Giulia Mascagni ⊕ Vedanth Nair ⊕ Mazhar Waseem ⊕ Miguel Almunia, “Does the Value-Added Tax Add Value? Lessons Using Administrative Data from a Diverse Set of Countries” . . . . . 107

**Articles**

Andrew Metrick, “The Failure of Silicon Valley Bank and the Panic of 2023” . . 133

Zoë Cullen, “Is Pay Transparency Good?” . . . . . 153

Olivier Marie and Paolo Pinotti, “Immigration and Crime: An International Perspective” . . . . . 181

Nancy Folbre, “Care Provision and the Boundaries of Production” . . . . . 201

Eliana Carranza and David McKenzie, “Job Training and Job Search Assistance Policies in Developing Countries” . . . . . 221

**Features**

Timothy Taylor, “Recommendations for Further Reading” . . . . . 245

## **Statement of Purpose**

The *Journal of Economic Perspectives* aims to bridge the gap between the general interest business and financial press and standard academic journals of economics. 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**

Kerwin Charles, Yale University

Cecilia Conrad, Lever for Change

Gary Hoover, Tulane University

Jeffrey Kling, Congressional Budget Office

Dylan Matthews, Vox.com

Emily Oster, Brown University

Catherine Rampell, Washington Post

Otis Reid, Open Philanthropy

Ted Rosenbaum, Federal Trade Commission

Lucie Schmidt, Smith College

David Wessel, Brookings Institution

# Skilled Immigrants, Firms, and the Global Geography of Innovation

Britta Glennon

**H**ighly skilled workers are a crucial and relatively scarce input into the productive and innovative processes of firms. The relevant talent pool for these workers is global. For example, 59 percent of artificial intelligence PhD-graduate hires by US firms are immigrants (Zwetsloot et al. 2019), indicating that the talent that fuels the growing artificial intelligence industry in the United States comes from across the globe.

As a result, many countries share a common objective of attracting and selecting the most talented immigrants who will contribute economically, who will possess valuable skills, and who will integrate both socially and economically. However, nations have taken very different approaches toward this objective, both in terms of the system by which skilled immigrants are selected and in terms of the relative emphasis placed on skilled immigration relative to other types of immigration.

The United States is one of the top destinations for immigrants; in 2020, 18 percent of the world's immigrants went to the United States (McAuliffe and Khadria 2020). The US immigration system permits legal immigration through three main pathways: family ties, humanitarian protection, and employer ties. The family ties pathway is the most common—roughly two-thirds of legal immigration to the United States is on the basis of family ties according to the Migration Policy

■ *Britta Glennon is an Assistant Professor of Management, Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania. She is also a Faculty Research Fellow, National Bureau of Economic Research, Cambridge, Massachusetts. Her email address is [bglennon@wharton.upenn.edu](mailto:bglennon@wharton.upenn.edu).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.1.3>.

Institute, although this varies each year. But my focus here is on *skilled* immigrants. For this group, the United States operates under a demand-driven system, in which US firms are the most important actors. Firms select the skilled immigrants that they wish to hire, submit the visa applications for these immigrants, and continue to sponsor them until the immigrant exits the country or becomes a permanent resident.

However, the demand from US firms for skilled immigrants exceeds the caps, rules, and national-level limits that apply. Many of the skilled immigrants that these firms want to hire started as international students at US universities, yet one estimate from Beine, Peri, and Raux (2023) suggests that only 23 percent of these international students in master's programs (8 percent of international students in bachelor's programs) transition into the US workforce because of these constraints. As an example, employment-based green card demand from employers currently exceeds the per-country limit for individuals from China, India, Mexico, and the Philippines, adding to their wait time for permanent residency (<https://travel.state.gov/content/travel/en/legal/visa-law0/visa-bulletin/2023/visa-bulletin-for-april-2023.html>). For citizens of India wishing to work in the United States, these wait times are currently estimated to exceed 100 years. Such wait times stand in sharp contrast to, for example, green cards based on immediate family ties, which face none of these caps. These issues often surface in particular contexts; for example, a major challenge for semiconductor firms seeking to take advantage of the incentive in the Creating Helpful Incentives to Produce Semiconductors and Science Act of 2022 (the "CHIPS Act") has been accessing the relevant human capital, given tight immigration restrictions.

This essay begins with an overview on the policy environment in the United States and abroad for skilled immigration, including "supply-based" systems where immigration is determined by a scale that gives points for various education, skill, and financial characteristic, and "demand-based systems" like the United States that primarily run through employers. I review the US system in some additional detail.

I then examine the evidence on how skilled immigrants and skilled immigration policy influence firm decision-making and firm outcomes. In response to changes in the availability of skilled immigrants, firms can change the structure of their production function, change their scope, or move the employment of skilled human capital abroad. Hence, skilled immigration policy influences firm decision-making on the location of where they locate their skilled activities, how they conduct those activities, and their broader human resources strategy. Skilled immigrants also have a significant impact on various dimensions of firm performance, including foreign investment, innovation, and survival, all of which are likely to affect economic growth, national competitiveness, and long-term labor market consequences for job creation, job stability, and earnings. Taking these issues a step further, the policies that affect skilled migration will affect the global geography and quality of innovation. In the final section, I discuss policy implications and open questions.

## The Role of Firms in Skilled Immigration Policy

Countries have taken two main approaches to skilled immigration policy, with very different implications for the role of firms. Here, I discuss the merits of supply- and demand-based policies for skilled immigration, highlighting the active role of firms in both processes. I then describe in more detail the demand-driven US process for skilled immigration and the types of firms that apply for skilled immigration visas in the United States.

### Skilled Immigration Policies across Countries: A Comparative Overview

Cross-country variations in skilled immigrant policy can be broadly categorized into two general approaches: some recipient countries implement supply-driven policies, such as the points systems observed in Canada, New Zealand, and Australia, while others adopt demand-driven policies, such as the employer-based admissions system in the United States and some European countries. Table 1 provides an overview of the key features of each system. A more detailed review can be found in Papademetriou and Hooper (2019) and Aydemir (2020), while an analysis of the effectiveness of these policies can be found in Czaika and Parsons (2017).

Points-based systems—the most common form of supply-driven policies—are inherently top-down in nature. In such systems, the government determines a set of desirable characteristics for immigrants, assigns points to these characteristics, and establishes a threshold number of total points. Prospective immigrants who exceed the threshold are then either granted admission or prioritized for permanent residency. Some countries with point systems, like the United Kingdom and Denmark, do not grant permanent residency right away, but instead use a points-based system to admit workers on an initially temporary basis. Desirable characteristics that are rewarded in a points system frequently include language proficiency, education, age, and either a job offer or work experience; however, the emphasis on different characteristics varies considerably between countries. For instance, Canada's point system prioritizes language proficiency and education, while Australia places the highest points on age. In the past, points-based systems have been largely separate from whether a skilled immigrant had already secured employment, but most countries that use a points-based system have added some form of employer-selected immigration (or expanded the points for such an employment offer) to their system fairly recently. For example, Canada and New Zealand award points for job offers, and Australia has expanded the share of immigrants who enter through employer-selected routes.

To provide a concrete example, Canada's Comprehensive Ranking System currently allocates points along four main dimensions: (1) Core/Human Capital Factors (worth up to 460 points without a spouse/partner and up to 500 points with a spouse/partner); (2) Spouse or Common-Law Factors (worth up to 40 points); (3) Skill Transferability Factors (up to 100 points); and (4) Additional Points (up to 600 points). Each person who submits a profile is assigned a score out of 1,200 total points.

Table 1

### A Comparative Overview of Supply-Driven versus Demand-Driven Skilled Immigration Policies

	<i>Supply-driven (Points-based)</i>	<i>Demand-driven (Employment-based)</i>
Selection actor	Governments, through points allocation	Firms, through job offers
Value placed on	Long-term desirable human capital characteristics	Short-term labor demand by firms
Role of government	Setting the quantity Setting the selection criteria	Setting the quantity Setting the regulatory parameters for selection
Role of firms	Minimal	Selection and sponsorship of immigrants
Primary advantages	Transparent and flexible. Freedom of movement between employers.	Responds directly to the needs of firms. Ensures economic integration.
Primary disadvantages	No guarantee that immigrants will find work easily. Concern about “brain waste” Selection on observables	Sets up the possibility of worker exploitation. Employers can manipulate the system to access cheaper labor. Moral hazard problem. Not as effective in addressing longer-term needs
Example countries	Canada, Australia, New Zealand	United States, some European countries, Japan, Korea

In the Canadian system, “Core/Human Capital Factors” includes age (where the top scoring age category is 20–29), level of education, English/French proficiency, and Canadian work experience, with the highest point category being English/French proficiency and the lowest point category being Canadian work experience. The points also vary depending on whether the applicant has a spouse or common-law partner. “Spouse or Common-Law Factors” mirror the Core/Human Capital Factors but are for the applicant’s spouse/partner. “Skill Transferability Factors” weights education by whether the degree was in English or French and also considers foreign vs. Canadian work experience. “Additional points” considers whether the candidate has a sibling living in Canada (15 points), speaks French (50 points), has a post-secondary education in Canada (30 points), has an employment offer (200 points), or has a nomination from a particular province (600 points). The Canadian government then conducts a draw every two or so weeks and issues a round of invitations to apply for permanent residence to the highest-ranking candidates. The minimum required score and the number of immigration invitations extended are not disclosed and may vary substantially by draw.

Demand-driven policies for immigrant skilled labor, in contrast, are characterized by the requirement of a job offer for admission. Governments regulate the pool of human capital from which firms can select, through criteria such as the minimum level of education, minimum paid wages, requirements that employers search for a native worker before hiring a skilled immigrant, and fees per foreign worker hired, but it is the firm who selects and sponsors the immigrant. Because



the firm—not the immigrant—submits the visa application, the immigrant’s right to live and work in the host country is typically tied to a specific employer. Under employment-based systems, immigrant workers are typically provided temporary visas first, but can apply for permanent residency after a few years.

The United States is the leading example of a country where demand-driven policies are used to select skilled immigrants and will be described in more detail below. Outside the United States, many of the skilled immigrants in Europe, Japan, and Korea are admitted through systems with elements of the demand-driven model. For example, to be eligible for Japan’s Highly Skilled Professional Visa, an applicant must be sponsored by a company *and* reach at least 70 points on a points-based system.

Each policy category has its advantages and disadvantages. Points-based systems are transparent and flexible, and they are better suited to admitting large flows of high-skilled immigrants. However, they can suffer from what some critics call “brain waste.” Because employers are traditionally not involved, and because it is difficult to evaluate characteristics such as motivation and creativity, points-based systems often admit immigrants who are unable to find work at their skill level once they arrive. As a result, many countries with points-based systems have moved towards including some elements of employment-based systems to ensure they admit the types of immigrants that firms want.

Employment-based systems, in contrast, by placing the firm in the active selection role, ensure that immigrants will have a job when they arrive and that their skills and credentials are valued in the labor market. This job connection improves economic integration. However, employment-based systems also have significant downsides. Critics often charge that employers can manipulate these systems more easily to access cheaper labor (for example, Hira 2018). Because the visas are tied to the employer, the system also sets up the possibility of exploitation of workers by firms: workers cannot easily change jobs or stand up to exploitative employers. If they are laid off, they also lose their legal immigration status. In contrast, immigrants admitted under points-based systems are able to move freely between employers.

Finally, there are other, less common pathways by which skilled immigrants can enter some countries, such as a “start-up visa” or an “investment visa.” An overview of start-up visas—which do not currently exist in the United States but have become increasingly popular around the world in the past decade—can be found in Kerr and Kerr (2021). Start-up visas typically seek to attract a particular type of skilled immigrant—entrepreneurs—who meet a particular set of requirements, typically around at least one of the following criteria: the entrepreneur’s qualifications and/or self-sufficiency, the estimated quality/impact of the start-up or start-up idea, and how new the start-up is. However, these start-up visas also vary quite substantially by country in terms of eligibility requirements, benefits, and the application process. For instance, the Start-Up Chile program provides a temporary visa without a path to permanent residency, while the Canada Start-Up Visa program provides permanent residency. However, Chile’s program provides an equity-free grant and a complete networking and mentoring program, while recipients of Canada’s Start-Up Visa

program are not provided anything beyond permanent residency by the government. In contrast with the growing popularity of start-up visas, investment visas, which provide a visa in return for making qualifying investments, have been phased out in many countries, such as Canada and Ireland, although they continue to exist in the United States.

In addition to differences in the approach for selecting skilled immigrant workers, countries vary significantly in the degree to which they emphasize the entry of skilled workers relative to other types of immigrants. For example, the United States places a relatively high weight on family-based visas (45 percent of all immigrants in 2011 were family-based) relative to work-related visas (9 percent in 2011), while Britain does the reverse (family-based immigrants made up only 8 percent of all immigrants in 2011 while work-related immigrants made up 35 percent). There are also large differences in per capita skilled immigrants permitted per country. For instance, work-related immigrants accounted for 0.04 percent of the United States' total population in 2011 but 0.51 percent of Australia's total population. A deeper dive into these immigration policy differences can be found in Boucher and Gest (2018), which is also the source for the statistics listed in this paragraph.

### **Skilled Immigration Policy in the United States: An Overview**

Skilled immigrants can come to the United States on either a temporary or permanent basis, and in both cases, the firm plays an active role. There are over 20 types of temporary employer-based visas, with variations in eligibility, duration, ability to bring dependents, ability to transition into permanent residence, and likelihood of success. The primary categories for skilled work visas are described in more detail in Table 2 and below.

The H-1B visa is the most common for skilled immigration: in fiscal year 2022, it was granted to about twice as many people as the other visa categories combined. It is available to “specialty occupations,” defined by the US Department of Labor as one that “requires the application of a body of highly specialized knowledge and the attainment of at least a bachelor’s degree or its equivalent.” The definition clearly gives considerable latitude to employers who wish to apply on behalf of an employee. The annual cap on the number of H-1B visas that can be issued is 85,000 H-1B visas, with 20,000 prioritized for holders of graduate degrees. The cap does not apply to visas granted to nonprofits—which includes most educational institutions—and also does not apply to renewals (these visas can be renewed for up to two three-year periods). The differences in cap constraints across sectors has altered the occupational choices of foreign students, leading them to substitute towards nonprofits and away from private sector firms (Amuedo-Dorantes and Furtado 2019). Cap-subject H-1B petitions are distributed in a first-come, first-served fashion in low-demand years, or by lottery in years where the number of petitions exceed the cap. A lottery has been needed in every year since 2004, when the cap was reduced from its previous high-mark of 195,000. Since 2014, the gap between supply and demand has become especially and consistently large; the annual number of cap-subject H-1B petitions requested has typically been two to three times greater than the cap.

*Table 2*  
**Temporary Skilled Worker Visas in the United States**

	<i>H-1B</i>	<i>L</i>	<i>O</i>	<i>TN</i>
Dual Intent?	Yes	Yes	Yes	No
Cap	85,000 (with 20,000 prioritized for graduate degree holders). Renewals and nonprofits do not count against the cap.	None	None	None
Who qualifies	“Specialty occupations”	Employees of multinational companies. The employee must have worked abroad at the multinational corporation for one continuous year in the last three years in a managerial/executive position (L-1A) or in a specialized knowledge staff capacity (L-1B).	Workers of “extraordinary ability”	Canadian and Mexican citizens who have a job offer with a firm and are in particular occupations
Length of time	Up to two three-year work periods (longer for those waiting for green cards)	L-1A: three years, extendable up to seven years L-1B: two years, extendable by two years	Three years, annually renewable	Unlimited (but cannot transition to PR)
Can spouses work in the US?	Originally no, but yes since 2015	Yes	Yes	No
Rejection Rate (FY2022)	42% selection rate from lotteries  2% rejection rate after lottery selection	16.4%	5.4%	8.6%
Number of visas issued (FY2022)	206,002	72,448	19,102	2,376

*Source:* The rejection rates listed in the table come from USCIS Nonimmigrant Worker Petitions by Case Status and Request for Evidence (RFE) ([https://www.uscis.gov/tools/reports-and-studies/immigration-and-citizenship-data?ddt\\_mon=&ddt\\_yr=&query=RFE&items\\_per\\_page=10](https://www.uscis.gov/tools/reports-and-studies/immigration-and-citizenship-data?ddt_mon=&ddt_yr=&query=RFE&items_per_page=10)). Additional details on the rejection rates for H-1B visas are from <https://www.uscis.gov/working-in-the-united-states/temporary-workers/h-1b-specialty-occupations-and-fashion-models/h-1b-electronic-registration-process>. The counts of number of visas issued in FY2022 are from State Department Table XV(B) Nonimmigrant Visas Issued by Classification (<https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/nonimmigrant-visa-statistics.html>).

*Note:* Issued numbers do not include spouses or children. The COVID-19 pandemic had an impact on the provision of visa-related services, such that fewer visas than usual were issued during the pandemic, so these numbers are in some cases higher than had been typical prior to the pandemic. The meaning and sources of other terms are discussed in the text.

In FY2022, the selection rate from the lottery was 42 percent, but nearly all petitions that made it through the lottery were approved.

The second most common is an L visa, for managers/executives or specialized knowledge staff who were already working at multinational companies in another country. O-1 visas, in turn, are for workers of “extraordinary ability,” defined by the

US Customs and Immigration services as individuals who possess “extraordinary ability in the sciences, arts, education, business, or athletics, or who has a demonstrated record of extraordinary achievement in the motion picture or television industry and has been recognized nationally or internationally for those achievements” (for more detail, see <https://www.uscis.gov/working-in-the-united-states/temporary-workers/o-1-visa-individuals-with-extraordinary-ability-or-achievement>). Finally, the TN visa, created by the North American Free Trade Agreement (NAFTA), is for Canadian and Mexican citizens who have a job offer from a firm and fall into certain occupational groups. There is no cap on any of these visas, but they also have higher rejection rates than the H-1B visa; they are more frequently denied by US Customs and Immigration. In particular, the L-1 visa has faced rejection rates ranging from 15 to 30 percent since 2015.

All of the visas listed in Table 2 are employer-based and temporary, which means that immigrants under these visas must work for the employer that petitioned for them, with limited ability to change jobs. If these workers are laid off or if their status expires, they typically must leave the United States. The first three visas have “dual intent,” which means that they permit a transition from temporary to permanent immigration, with an employer-sponsored “green card,” which is given to those with authorization to live and work in the United States on a permanent basis. Immigrants remain bound by the terms of their temporary visa until that point. The first three also allow spouses to come work in the United States. In contrast, the TN visa has neither of these provisions.

Two important visa categories are excluded from Table 2 because they are not strictly employment-based skilled work visas, but they are nonetheless programs that skilled immigrants can use to work in the United States. The first is the Optional Practical Training or OPT program. Unlike the other visas described above, the OPT program is not a visa, but instead allows temporary employment for international students in the United States under an F-1 student visa. Only foreign students on an F-1 student visa with a higher education degree from a US college or university are eligible for this program, and foreign students do not require employer sponsorship to apply for it. However, the program is time-limited. The standard program allows up to twelve months of OPT employment, but the program is longer for students in science, technology, engineering, and mathematics fields. In 2008, an extension was granted for up to 29 months of work authorization for F-1 nonimmigrant students with these degrees, and in 2016, it was extended even further to 36 months. The OPT program is a popular pathway that foreign students on F-1 visas use to remain in the United States for longer, in particular because it gives students more time to apply multiple times for employer-sponsored H-1B visas. Indeed, Beine, Peri, and Raux (2023) found that the 2008 extension increased transition rates to the US workforce for master’s graduates of programs in science, technology, engineering, and mathematics by 6.5 percentage points. The second is the J visa, which is an exchange visitor nonimmigrant visa. This visa permits a wide range of work-and-study-based exchange visitor programs, spanning from an au pair, camp counselor, or summer intern to a scientific researcher or professor.

The duration of the visa depends on the particular program. Importantly, the J visa is a nonimmigrant intent visa; it cannot be converted to a green card.

For those who wish to remain in the United States permanently, the next step after a temporary visa is to obtain a Permanent Resident Card, commonly known as a “green card.” The US government distributes roughly one million green cards each year in recent years. Most of them go to immediate family members of US citizen; for example, parents, spouses, or children under the age of 21. However, green cards are also available for skilled immigrants to transition into permanent residency through employment sponsorship. The cap on permanent employment-based immigrants is 140,000 per year, which constitutes about 14 percent of all green cards. This limit includes the immigrants as well as their eligible spouses and children. Apart from the numerical limit placed on the number of employment-based green cards, there is also a limit on the number of green cards that can be granted to individuals from any one country: no country can receive more than 7 percent of the total number of employment-based and family-sponsored green cards in a given year. As a result, citizens from countries with large populations—such as India or China—are typically subject to exceedingly long wait times for a green card. As noted earlier, these wait times currently are estimated to exceed 100 years for citizens of India.

Given the importance of the firm in US skilled immigration policy, it is useful to understand some basic characteristics of what types of firms sponsor skilled employment visas and how they differ from the modal US firm. Although detailed data are not readily available for most of the skilled visa categories, the H-1B Employer Data Hub managed by the US Citizenship and Immigration Services makes it possible to see some basic information about the firms who successfully petition for H-1B visas, which as shown in Table 2, made up about two-thirds of issued skilled employment-based temporary visas in fiscal year 2022.

A relatively small group of industries accounts for most of the H-1B visas.<sup>1</sup> Figure 1 illustrates that most firms utilizing the program are in the same industry—Professional, Scientific, and Technical Services—with Manufacturing a distant second. Indeed, in 2022, just five firms (Tata Consultancy, Amazon, Google, Microsoft, Infosys) accounted for nearly 10 percent of all H-1B approvals.<sup>2</sup> This skewness has increased over time (Kerr, Kerr, and Lincoln 2015; Mayda et al. 2020). The skewness is apparent geographically as well; 22.5 percent of approved H-1B petitions were allocated to employers in California alone.

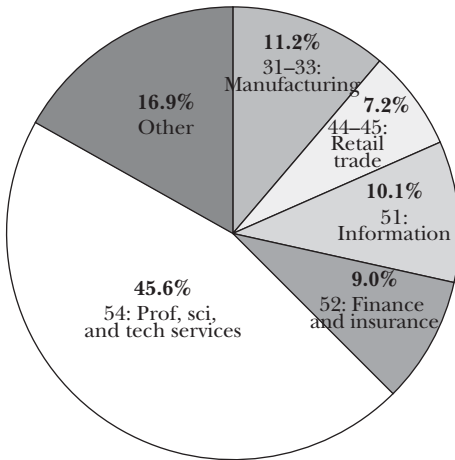
The H-1B Employer Data Hub does not provide any detail on the characteristics of H-1B employers beyond their industry and geographic location. Fortunately,

<sup>1</sup>I am, of course, not the first to document this pattern. A comprehensive overview of facts and trends, including the increase in concentration, using the H-1B data, can be found in Mayda et al. (2020) and Mayda et al. (2023). Kerr, Kerr, and Lincoln (2015) document some similar facts using data from the Labor Condition Application that employers fill out when applying for these visas.

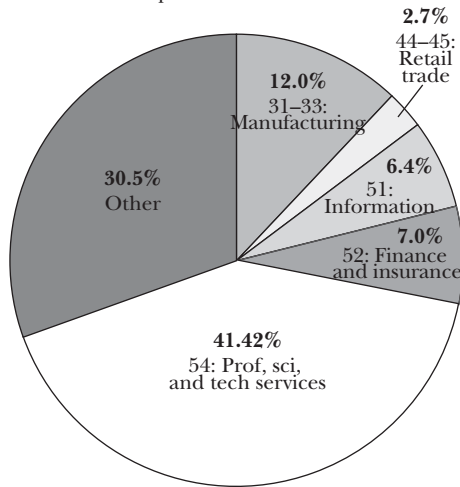
<sup>2</sup>If restricting the search to new petitions only, the top five companies change slightly (Amazon, Infosys, Tata Consultancy, Cognizant, and Google) but the concentration remains—these top five companies account for nearly 10 percent of all new and continuing H-1B approvals.

*Figure 1*  
**Industry Concentration of H-1B Employers**

Panel A. Industry distribution of H-1B issued petitions in FY2022



Panel B. Industry distribution of firms with  $\geq 1$  H-1B issued petition in FY2022



Source: USCIS H-1B Employer Data Hub (<https://www.uscis.gov/tools/reports-and-studies/h-1b-employer-data-hub>). Underlying data are simple aggregated counts over industry.

Note: Panel A indicates the share of total H-1B petitions in fiscal year 2022 in each industry category. Nearly half of all H-1B petitions in that year were in Professional, Scientific, and Technical Services. Panel B indicates the share of firms with at least one H-1B petition in fiscal year 2022 in each industry category.

other authors have utilized linkages between H-1B data and other datasets to produce some basic facts about how H-1B employers differ from other firms. For instance, Mayda et al. (2020) show, using a Compustat linkage, that H-1B employers spend 37 times more on research and development expenditures and have 5.7 times more employees than the average non-H-1B employer. H-1B employers also appear to be higher-growth companies. Doran, Gelber, and Isen (2022), using a linkage to IRS data, similarly find that H-1B firms are larger, more profitable, and have more patents.

## How Skilled Immigration Affects Firm Decision-Making and Outcomes

A standard concern about immigration, skilled or otherwise, is that it will take jobs from native-born workers. However, a nascent literature suggests that firms have many other margins of adjustment when their access to skilled immigrants is affected by national immigration policy. I first survey the now ample evidence

that skilled immigrants have a strong positive effect on firm (particularly start-up) performance, innovation, and investment to highlight why firms are likely to need to adjust their strategies when access to immigrants is limited. I then describe the potential margins of adjustment. In the next section, I discuss the implications of constrained access to skilled immigrants, given their importance to firms, for the global geography and quality of innovation.

### **How High-Skilled Immigration Affects Firm Outcomes**

A growing literature has examined how access to skilled immigrants can affect firm outcomes, including innovation, performance, access to venture capital funding, and foreign investment and trade. Understanding whether access to skilled immigrants affects firm outcomes lays the groundwork for understanding why firms are likely to leverage many response margins when this access is severely constrained (as is often the case in places like the United States, as shown in the previous section). I focus first on firm outcomes within national borders, and then across national borders.

*Effect on firm outcomes within borders.* A burgeoning literature has linked high-skilled immigration to various dimensions of innovation. Some of this work has been done at the individual level; for example, documenting the higher propensity of immigrants to patent (Bernstein et al. 2022; Hunt 2011; Hunt and Gauthier-Loiselle 2010) or the positive impact of skilled immigrants on the innovativeness of their peers (Borjas and Doran 2012; Ganguli 2015; Hunt and Gauthier-Loiselle 2010; Moser, Voena, and Waldinger 2014). Other work has been done at the regional level, like the positive effect of skilled immigrants on local innovation (Akcigit, Grigsby, and Nicholas 2017; Burchardi et al. 2020; Kerr and Lincoln 2010). Less work, until recently, has examined whether these linkages translate to the firm setting: does hiring more skilled immigrants improve firm-level innovation, productivity, and performance?

It is not immediately apparent that, despite immigrant-innovation linkages at the individual and regional levels, hiring more skilled immigrants would lead to positive effects on firm innovation. For example, if skilled immigrants and natives are perfect substitutes and skilled immigrants simply replace the skilled natives within the firm, then the outcomes for the firm will remain the same—indeed, this explanation has been proposed for the null effect of H-1Bs on firm patenting found in Doran, Gelber, and Isen (2022). Hiring skilled immigrants could even have a negative impact on performance if an increase in the immigrant workforce creates conflict and coordination frictions that lower performance or if it reduces knowledge transfer among workers due to language and cultural barriers (Page 2011; Parrotta; Pozzoli, and Pytlikova 2014).

On the other hand, skilled immigrants might improve firm-level innovation and productivity for a variety of reasons. First, if there are shortages of skilled workers, then greater access to skilled immigrants allows firms to increase their stock of human capital at the firm to the optimal level, thus stimulating innovation and productivity, particularly among those firms that were facing skilled labor shortages.

Some empirical evidence supports this channel. For example, Beerli et al. (2021) look at labor reforms which allowed more skilled workers to enter Switzerland, and found an increase in the size, productivity, and innovation of these firms, leading to an increase in wages for native-born Swiss skilled workers. Brinatti et al. (2023) find that skill-intensive and high-paying US firms that won the H-1B lottery in 2007 were more likely to expand their scale and productivity, with limited evidence of any negative effects for skilled native-born workers. Ghosh, Mayda, and Ortega (2014) study the effects of a sharp reduction in the gap for H-1B visas in 2004 and find that relaxing the cap would allow a subgroup of firms with intensive research and development to achieve higher scale and productivity. This explanation could explain firm productivity gains to hiring more skilled immigrants regardless of whether such immigrants are the same as or different from skilled natives.

A second explanation, in contrast, posits that skilled immigrants are unique, with different talents, and this difference leads to improved firm innovation. Countries specialize in different technology areas; there is no global pool of technology (Keller 2004). As a result, immigrants are likely to bring a different set of knowledge than their native counterparts. As an illustration of this pattern, Bahar, Choudhury, and Rapoport (2020) find that when migrant inventors come from a country that specializes in certain technologies, they are more likely to patent in the same technology areas in the receiving country. Because the creation of new technical knowledge often involves the novel recombination of known knowledge elements (Fleming 2001; Henderson and Clark 1990), the combination of the knowledge of natives and that of different immigrants could lead to more innovation. Indeed, immigrants can increase the variety of novel ideas and know-how, thus allowing firms to innovate and engage in a greater variety of actions themselves, as illustrated by Choudhury and Kim (2019) in the context of herbal patents; Glennon et al. (2021) in the context of European football; and Moser, Voena, and Waldinger (2014) in the context of Jewish emigres fleeing Nazi Germany.

Third, skilled immigrants may also differ in being generally more talented, having been drawn from the right tail of the talent distribution because of the frictions involved in moving across borders (Borjas 1987; Glennon et al. 2021), or specialized according to their comparative advantages (Mayda, Orefice, and Santoni 2022; Peri and Sparber 2009). Finally, as will be discussed more in the next subsection, skilled immigrants could improve firm productivity through either prompting a change in adopted production techniques or through complementarities between skilled immigrants and technology.

The discussion above is restricted to the effects of skilled immigration on existing, incumbent firms; while the empirical evidence thus far lands mostly on the side of skilled immigrants having a strong positive impact on firm innovation, productivity, and performance, there is still some debate.<sup>3</sup> The effects of skilled

<sup>3</sup>One important study has found zero impact of hiring skilled immigrants on firm innovation (Doran, Gelber, and Isen 2022). A direct comparison of the Doran, Gelber, and Isen (2022) study with the Brinatti et al. (2023) study—both of which use the same setting (US firms) and identification strategy



immigrants on start-up performance—along a host of metrics including start-up creation, survival, success, and ability to attract venture capital funding—are, in contrast, more clear-cut. It may seem unlikely that immigrants will make a strong contribution to start-ups in the US economy, given that its skilled immigration policies tend to suppress entrepreneurship as an early career choice of immigrants (Agarwal, Ganco, and Raffiee 2021). However, immigrants are more likely than natives to start firms of all sizes, as Azoulay et al. (2022) show by drawing upon a range of data from administrative data on US firms to the Fortune 500 (see also Fairlie and Lofstrum 2015; Kerr and Kerr 2020). In addition, start-ups who can hire H-1B workers are more likely to have a successful exit via an initial public offering in the stock market or acquisition, and are more likely to receive additional venture capital funding (Dimmock, Huang, and Weisbenner 2022; Chen, Hshieh, and Zhang 2021).<sup>4</sup> Some of that additional venture capital funding comes from abroad; hiring skilled immigrant workers improves start-ups' access to international venture capital funding because of their home-country professional networks and familiarity with the home country culture (Li 2020).

*Effect on firm outcomes across borders.* Immigrants may play an even larger role for firm outcomes across borders. There is a well-documented strong positive relationship in the research literature going back several decades between immigrant diasporas and the trade, capital, and investment linkages with the countries of origin of these immigrant diasporas. For some recent examples (and of course, interested readers will want to check earlier studies cited in these works), Burchardi, Chaney, and Hassan (2019) find that the presence of historical diasporas in US counties leads to more foreign direct investment to and from those counties and the countries of origin of the diasporas, and Balachandran and Hernandez (2021) find that US venture capital firms become more likely to invest in India as their ties expand to Indian entrepreneurs in the United States. Hernandez (2014) further links the presence of immigrant diasporas to the success of foreign subsidiaries.

These immigrant diasporas increase cross-border investment success through three main channels. First, immigrants can reduce information frictions; in particular, because immigrants understand features of their home country such as market structure, consumer preferences, and business norms, they can identify business

---

(the H-1B lottery)—reveals that firm heterogeneity is likely at least part of the explanation. Brinnatti et al. (2023) write, “The main reason that these results diverge is likely that the sample of firms that applied for visas on the last day of filing for 2005 and 2006 (used by DGI) are different than the full sample of H-1B applicants that applied in 2007 (as in our paper). As shown in Section 7, firms that expand the most after winning the lottery are high-paying firms that are skill-intensive and immigrant-intensive. Hence, firms that really need the immigrants for production are likely to submit their applications as soon as they can, while firms where immigrants are not as critical to production may apply with less urgency.” A likely implication is that the positive innovation and productivity benefits from an improvement in access to skilled immigrant workers, but such benefits mainly exist for the subset of firms who are heavily reliant on such workers.

<sup>4</sup> In addition to facing visa cap constraints, Diethorn (2023) and Roach and Skrentny (2019) have found that skilled immigrants are already less likely to opt into working for a US start-up because of visa concerns.

opportunities and alleviate many of the challenges inherent in cross-border trade and investment. Second, diasporas increase demand for “nostalgia” goods from the country of origin and in that way influence consumer preferences and tastes. Third, diasporas can provide social capital; for example, when courts do not work well, the social connections of immigrants and their knowledge of the local regime can decrease the costs of enforcing contracts.

Hiring skilled immigrants can also affect multinational firm cross-border success. For example, Foley and Kerr (2013) find that US-based ethnic inventor employees helped US multinational corporations to form new affiliates abroad without needing the support of a local venture capital partner. These ethnic inventors also were associated with an increase in multinational innovation, particularly in countries related to that ethnicity. As with the literature linking immigrants to trade and foreign direct investment, Foley and Kerr (2013) argue that the ability of these ethnic inventors to facilitate these changes is likely due to their knowledge and understanding of origin country customer demand, how to conduct business in that country, personal connections, and the trust that a shared language and culture facilitates.<sup>5</sup> Kerr and Kerr (2018) extend this analysis to show US multinational corporations have more patents with cross-border teams—which account for much of the growth in global invention—when they have more US-based ethnic researcher employees. These papers suggest that skilled immigrants within multinational companies enhance the innovativeness and competitiveness of these firms abroad as well.

### **Margins of Adjustment When Immigration Is a Constraint**

Constraints on the ability to hire skilled immigrants can influence the decision-making of firms with respect to where they locate skilled activities, how they conduct those activities, and their broader human resources strategy. For each possible response, I describe the (often limited) empirical evidence that we have on these responses, divided into within-US adjustments and cross-border adjustments.

*Within-US adjustments.* First, policy changes affecting the level of skilled immigration could provide incentives for a firm to change its production technology. Although little is known about this response in the skilled immigration context, some papers have analyzed adjustments in production technologies in response to a change in unskilled immigration. For example, Clemens, Lewis, and Postel (2018) show that restrictions on low-skill immigration—specifically, the exclusion of Mexican bracero farm workers—induced firms to either adopt more capital-intensive production techniques or to change production levels as a substitute for the foreign-born labor they could no longer hire. Although the substitution of technology for labor, once labor is restricted, could be specific to low-skilled labor, the emergence of artificial intelligence technologies like ChatGPT suggests

<sup>5</sup> For other work on how skilled immigrants and ethnic ties facilitate knowledge transfer, useful starting points include Agrawal et al. (2011), Choudhury (2016), Fry (2023), Kerr (2008), MacGarvie (2006), Saxenian (2006), and Wang (2015).

that some high-skilled labor could also be susceptible (Eloundou et al. 2023; Felten, Raj, and Seamans 2023). More work is needed to understand whether firms might also respond to restrictions on skilled immigration with a heavier reliance on capital-intensive production technologies.

A separate strand of the literature, also focused on low-skilled immigration, has focused on capital-skill complementarities—that is, lower-skill immigration might be accompanied by less capital investment, and vice versa. In response to changes in low-skilled immigration in Germany, Dustmann and Glitz (2015) find evidence of these kinds of within-firm changes in production technology. Some research using US data has focused on immigration-induced variation in the skill mix: Doms and Lewis (2006) use establishment-level data to document that additional human capital led to more investment in personal computers; Lewis (2011) shows that manufacturing firms in US metropolitan areas that received a larger share of low-skilled immigration in the 1980s and 1990s were less likely to invest in capital equipment, which tended to complement middle-skilled labor; and Peri (2012) finds that states receiving a greater share of low-skill immigrants in the 1990s and 2000s led to a shift toward less capital-intensive production. These findings also suggest the possibility of the opposite: if less-skilled labor leads to less capital-intensive production, more-skilled labor could lead to more capital-intensive production.

Indeed, because skilled immigrants are themselves more innovative or entrepreneurial than natives—as discussed in the previous section—then immigration could speed technological progress, raising productivity and average wages. In short, firms may respond to changes in the flows of skilled immigrants by modifying their production functions in ways that can be difficult to foresee.

Second, firms may also change their scope; that is, they might change the set of functions they conduct in-house as opposed to those outsourced to other firms. When firms are unable to hire skilled immigrants themselves, they may instead contract with other firms that specialize in hiring and outsourcing skilled immigrant workers. In this arrangement, the skilled immigrants ultimately work for third-party clients. The H-1B cap and lottery system has led to a significant rise in this outsourcing behavior (Sharma and Sparber 2020), but the outsourcing model is likely to be a response that extends beyond the specific H-1B case. A related possible response available to firms, which also extends the boundaries of the firm, is more specific to the US skilled immigration regime. In the United States, nonprofit entities are exempt from the H-1B cap—specifically, this includes universities, nonprofit organizations affiliated with universities, nonprofit research institutions, and government research institutions. These institutions can apply for an H-1B visa at any time of the year and their pool of H-1B visas is not capped or subject to the lottery. Firms can therefore choose to partner with these cap-exempt entities, and by so doing tap into a pool of highly-skilled workers without being constrained by the H-1B visa cap. Anecdotally, organizations such as the Open Avenues Foundation—a nonprofit organization that partners with universities—exploit this arbitrage opportunity by matching cap-exempt employers with cap-subject employers. After Open Avenues Foundation hires a professional on a cap-exempt H-1B, a private

and cap-subject company can file a concurrent, also cap-exempt H-1B for the professional and therefore fill its H-1B position while avoiding the randomness of the lottery. On their website, this point is explicit: “Our programming offers a cap-exempt H-1B visa solution to skilled foreign talent” (see <https://www.openavenuesfoundation.org/>). Similarly, firms could create separate nonprofit entities (such as, for example, Microsoft Research) to hire skilled immigrants not subject to the H-1B visa cap. However, no empirical evidence of which I am aware measures the prevalence of this response margin.

*Cross-border adjustments.* Firms can also make adjustments to their geographic footprint across national borders. In particular, multinational firms are not constrained by national borders, so if they are unable to hire the skilled labor they need in one country, they can hire it elsewhere instead. Indeed, when US multinational companies were faced with restrictions on H-1B visas, they increased foreign affiliate employment in all areas, but especially in research and development at both the intensive margin (US multinationals employed more people at their existing foreign affiliates) and the extensive margin (US multinationals entered new countries) (Glennon 2023). The average multinational company created 0.42 foreign affiliate jobs at existing locations for every unfilled H-1B position, while the most internationalized multinational companies created 0.93 for every unfilled H-1B position, not including the opening of new foreign affiliates.

Similarly, early evidence suggests that some nonmultinational firms may respond to immigration restrictions by moving their location entirely; Lee and Glennon (2023) find that US immigrants change their start-up founding location to Canada after the introduction of Canada’s Start-Up Visa Program. But there is still relatively little understanding of the degree to which nonmultinational firms could respond to skilled immigration restrictions by either becoming multinational firms (opening a new foreign affiliate) or by moving their home country entirely.

Clearly, firms can respond to constraints on their ability to hire skilled immigrants in multiple ways. However, a deeper understanding of what types of firms use each response, how prevalent the responses are, and what interdependencies arise between the responses remains a task for future research. The next section builds on these results and discusses the possible implications for the geography of innovation.

## **How Skilled Migrants and Skilled Immigration Policy Shape Cross-Border Innovation**

In the previous section, I described two ways in which skilled immigrants can play an important role in the global location choices of multinational corporations, and perhaps other firms as well. First, hiring immigrants increases both inward and outward foreign direct investment and trade, expanding the global footprint of firms. Second, firms may respond to restrictive immigration policy by shifting employment and investment to other countries.

In addition to affecting cross-border outcomes and strategy, each of these channels, logically, is likely to shape the global geography of innovation. Hiring skilled immigrant inventors within multinational corporations facilitates the disintegration of innovative activity (like R&D and patents) across borders, expanding inventive activity abroad by these multinational corporations. This shift does not come at the cost of the home country, which still plays a central role.<sup>6</sup> In this way, the ability to hire skilled immigrants may allow multinational firms to *expand* their innovative activity, becoming not only be more innovative at home, but abroad as well, bringing positive-sum benefits to home and host countries.

However, the cross-border response of firms to restrictive immigration policy—whereby multinational firms unable to hire skilled immigrants at home hire them at their foreign affiliates instead—suggests that there is likely to be a *reallocation* of global innovative activities. Indeed, Bahar et al. (2022) analyze changes in multinational corporation patenting in response to business-related immigration reforms across 15 countries and finds that reforms discouraging inventor migration reduce both domestic and global collaborative patenting by multinational corporation subsidiaries. Furthermore, they find that business-related migration policies shifted the geography of innovation toward emerging markets. Restrictive immigration policies, therefore, are likely to shift innovative activities of multinational corporations away from countries with restrictive immigration policies and towards countries with less restrictive policies or towards the common countries of origin of immigrants.

Because the geographic location of skilled immigrants is linked to innovative spillovers (Hunt and Gauthier-Loiselle 2010; Moser, Voena, and Waldinger 2014), as is the geographic location of multinational innovative activity (for example, Javorcik 2004; Poole 2013), and because innovative spillovers are geographically localized, those location choices are likely to extend beyond the innovation done by the multinational corporation, potentially shaping host country innovative ecosystems as a whole. The second channel in particular—which shifts the location of innovation away from countries with restrictive skilled immigration systems like the United States—could have implications for the quality and quantity of innovation generally. For example, Agarwal et al. (2023) look at students who participated in the International Mathematical Olympiad and where they end up going to university and living. They find that those who go to the United States are far more productive, and offer a back-of-the-envelope calculation that “removing constraints on immigration could increase the global scientific output of future cohorts by up to 50 percent.” In a similar spirit, Kahn and MacGarvie (2016) compare foreign-born, US-educated scientists whose visas require them to return to their home country to those who do not, and find that researchers who can remain in countries with high per capita GDP have much higher research output than those who go to countries with low per

<sup>6</sup>There is a much larger literature on global collaboration and foreign R&D that is not touched upon here because it does not explicitly consider immigration: for an excellent overview, see Kerr and Kerr (2018).

capita GDP. These papers imply that restricting skilled immigration hurts not only the home country's national competitiveness by pushing innovation elsewhere, but also leads to big losses for science and innovation as a whole because that innovation may end up being lower quality.

The potential tradeoff, of course, is concerns that are raised about brain drain from lower-income countries. But the lasting economic connections between immigrants and their countries of origin emphasized in this essay suggest large potential benefits to immigrant-sending countries as well (for a more comprehensive overview of the brain drain arguments, interested readers should see Gibson and McKenzie 2011).

In short, innovative productivity can depend on location, and the willingness of the United States to accept more skilled immigrants can affect both US national competitiveness and the global advancement of science. More broadly, skilled immigration policy could potentially shape both the geography of innovative activity by multinational corporations and the innovation ecosystems and national competitiveness of countries around the world.

## **Discussion and Conclusions**

The vigorous public policy debate over immigration and the US economy has typically focused on the extent to which an increased supply of immigrant labor could have negative consequences for the wages or job opportunities of native-born workers. In recent years, researchers have explored these arguments in the context of models with workers of heterogeneous skill levels and variation in tasks, which thus allow for either substitutions or complementarities between immigrant and native-born labor (for an overview in this journal, see Peri 2016). In contrast, this essay has taken a resolutely demand-side approach to immigration. Firms play a central role in the skilled immigration process in most countries around the world, especially in demand-driven systems like the United States, but increasingly also in countries with points-based systems: companies select the skilled workers that they wish to hire.

The availability of skilled immigrants changes both the behavior of firms and the pattern of firm outcomes. Hiring skilled immigrants leads to positive performance effects with regard to both productivity and innovation for at least a significant subset of large firms, and they are critical for start-ups, both as founders and hires. They shape the foreign investment and trade decisions of firms, and positively affect firm success abroad. Skilled immigrants also shape *where* and *how* firms do innovation. In a global economy, firms located in countries with restrictive skilled immigration policies have a competitive disadvantage, particularly if they are not multinational companies.

This demand-side perspective argues that it is fundamentally mistaken to think of skilled immigrants as simply a replacement or as additional competition for native-born skilled labor. Instead, the disproportionate role of skilled immigrants

in starting firms means that skilled immigrants act as job creators, pushing out the supply curve. Positive effects on firm performance and investment similarly imply more hiring. Increases in innovation and technological progress due to immigrants can raise productivity and average wages. Restrictions on skilled immigrant hiring lead to more offshoring, which moves jobs from the home country abroad.

There is a need for richer immigration models that embed firm choices and dynamics with regard to changes in the availability of skilled immigration. Some research has begun to rise to this challenge, by incorporating firms and firm dynamics into models analyzing the welfare impact of skilled immigration (Mehra and Shen 2022; Waugh 2018), by including immigration and offshoring as interconnected decisions in general equilibrium models (Olney 2012; Ottaviano, Perry, and Wright 2013), or by accounting for the channel of multinational activity in welfare analysis (Morales 2019). More can and should be done to expand on this work.

In addition to incorporating firms into analysis of skilled immigration, there is also a need to take seriously the heterogeneity in firm behavior and outcomes with regard to skilled migration; for instance, ignoring heterogeneity in immigrant share across firms leads to an underestimation of welfare gains to native workers from immigration by 11 percent (Brinatti and Morales 2021). The role of multinational firms in particular has traditionally been underexplored, even though they are the largest hirers of skilled immigrants. In the US economy, US-based multinational companies account for about one-third of H-1B visa applications, not including foreign multinational companies who also hire H-1B workers (for instance, Morales 2019); not coincidentally, US multinational companies account for 80 percent of the research and development done by US firms. Multinational firms also play a disproportionate role in the circulation of human capital; inventors and senior managers within multinational companies are frequently transferred between different country offices. Similarly, we know very little about the differences in behavior and outcomes between nonprofit and for-profit firms with regard to immigration. The benefits that the trade literature gained from its emphasis on the role of the firm and the importance of firm heterogeneity, beginning with the work of Bernard and Jensen (1999) and Melitz (2003), suggest the potential importance of this approach to skilled migration.

Finally, evaluations of the impact of skilled immigration should not be constrained within borders: immigration flows and national immigration policies affect the global geography of innovation and investment. A change in skilled immigration in one country can have significant implications for innovation, investment, and employment in other countries. Differences in skilled migration policy may affect multinational firm structure, global business competition, and, especially, the global geography of innovation. Countries that hamstringing their own firms with restrictive skilled immigration policies may damage their own national competitiveness by shifting investment and innovation abroad.

## References

- Agrawal, Ajay, Devesh Kapur, John McHale, and Alexander Oettl.** 2011. "Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation." *Journal of Urban Economics* 69 (1): 43–55.
- Agarwal, Rajshree, Martin Ganco, and Joseph Raffiee.** 2021. "Immigrant Entrepreneurship: The Effect of Early Career Immigration Constraints on New Venture Formation." *Organization Science* 33 (4): 1372–95.
- Agarwal, Ruchir, Ina Ganguli, Patrick Gaulé, and Geoff Smith.** 2023. "Why U.S. Immigration Matters for the Global Advancement of Science." *Research Policy* 52 (1): 104659.
- Akcigit, Ufuk, John Grigsby, and Tom Nicholas.** 2017. "Immigration and the Rise of American Ingenuity." *American Economic Review* 107 (5): 327–31.
- Amuedo-Dorantes, Catalina, and Delia Furtado.** 2019. "Settling for Academia? H-1B Visas and the Career Choices of International Students in the United States." *Journal of Human Resources* 54 (2): 401–29.
- Aydemir, Abdurrahman B.** 2020. "Skill-Based Immigration, Economic Integration, and Economic Performance." IZA World of Labor 41.
- Azoulay, Pierre, Benjamin F. Jones, J. Daniel Kim, and Javier Miranda.** 2022. "Immigration and Entrepreneurship in the United States." *American Economic Review: Insights* 4 (1): 71–88.
- Bahar, Dany, Prithwiraj Choudhury, and Hillel Rapoport.** 2020. "Migrant Inventors and the Technological Advantage of Nations." *Research Policy* 49 (9): 103947.
- Bahar, Dany, Prithwiraj Choudhury, Sara Signorelli, and James Sappanfield.** 2022. "Talent Flows and the Geography of Knowledge Production: Causal Evidence from Multinational Firms." Harvard Business School Working Paper 22-047.
- Bahar, Dany, Ricardo Hausmann, and Cesar A. Hidalgo.** 2014. "Neighbors and the Evolution of the Comparative Advantage of Nations: Evidence of International Knowledge Diffusion?" *Journal of International Economics* 92 (1): 111–23.
- Balachandran, Sarath, and Exequiel Hernandez.** 2021. "Mi Casa Es Tu Casa: Immigrant Entrepreneurs as Pathways to Foreign Venture Capital Investments." *Strategic Management Journal* 42 (11): 2047–83.
- Beerli, Andreas, Jan Ruffner, Michael Siegenthaler, and Giovanni Peri.** 2021. "The Abolition of Immigration Restrictions and the Performance of Firms and Workers: Evidence from Switzerland." *American Economic Review* 111 (3): 976–1012.
- Beine, Michel, Giovanni Peri, and Morgan Raux.** 2023. "International College Students' Impact on the US Skilled Labor Supply." *Journal of Public Economics* 223: 104917.
- Bernard, Andrew B., and J. Bradford Jensen.** 1999. "Exceptional Exporter Performance: Cause, Effect, or Both?" *Journal of International Economics* 47 (1): 1–25.
- Bernstein, Shai, Rebecca Diamond, Abhisit Jiranaphawiboon, Timothy McQuade, and Beatriz Pousada.** 2022. "The Contribution of High-Skilled Immigrants to Innovation in the United States." NBER Working Paper 30797.
- Borjas, George J.** 1987. "Self-Selection and the Earnings of Immigrants." *American Economic Review* 77 (4): 531–53.
- Borjas, George J., and Kirk B. Doran.** 2012. "The Collapse of the Soviet Union and the Productivity of American Mathematicians." *Quarterly Journal of Economics* 127 (3): 1143–1203.
- Boucher, Anna K., and Justin Gest.** 2018. *Crossroads: Comparative Immigration Regimes in a World of Demographic Change*. Cambridge, UK: Cambridge University Press.
- Branstetter, Lee G.** 2001. "Are Knowledge Spillovers International or Intranational in Scope? Microeconomic Evidence from the U.S. and Japan." *Journal of International Economics* 53 (1): 53–79.
- Branstetter, Lee G., Britta Glennon, and J. Bradford Jensen.** 2018. "The IT Revolution and the Globalization of R&D." *Innovation Policy and the Economy* 19: 1–37.
- Branstetter, Lee, Guangwei Li, and Francisco Veloso.** 2015. "The Rise of International Co-invention." In *The Changing Frontier: Rethinking Science and Innovation Policy*, edited by Adam B. Jaffe and Benjamin F. Jones, 135–68. Chicago: University of Chicago Press.
- Brinatti, Agostina, Mingyu Chen, Parag Mahajan, Nicolas Morales, and Kevin Shih.** 2023. "The Impact of Immigration on Firms and Workers: Insights from the H-1B Lottery." Unpublished.
- Brinatti, Agostina, and Nicolas Morales.** 2021. "Firm Heterogeneity and the Impact of Immigration: Evidence from German Establishments." Federal Reserve Bank of Richmond Working Papers 21-16.



- Burchardi, Konrad B., Thomas Chaney, and Tarek A. Hassan.** 2019. "Migrants, Ancestors, and Foreign Investments." *Review of Economic Studies* 86 (4): 1448–86.
- Burchardi, Konrad B., Thomas Chaney, Tarek Alexander Hassan, Lisa Tarquinio, and Stephen J. Terry.** 2020. "Immigration, Innovation, and Growth." NBER Working Paper 27075.
- Chen, Jun, Shenje Hshieh, and Feng Zhang.** 2021. "The Role of High-Skilled Foreign Labor in Startup Performance: Evidence from Two Natural Experiments." *Journal of Financial Economics* 142 (1): 430–52.
- 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* 16 (3): 585–610.
- Choudhury, Prithwiraj, and Do Yoon Kim.** 2019. "The Ethnic Migrant Inventor Effect: Codification and Recombination of Knowledge across Borders." *Strategic Management Journal* 40 (2): 203–29.
- Clemens, Michael A., Ethan G. Lewis, and Hannah M. Postel.** 2018. "Immigration Restrictions as Active Labor Market Policy: Evidence from the Mexican Bracero Exclusion." *American Economic Review* 108 (6): 1468–87.
- Czaika, Mathias and Christopher R. Parsons.** 2017. "The Gravity of High-Skilled Migration Policies." *Demography* 54 (2): 603–30.
- Diethorn, Holden A.** 2023. "Green Card Quotas and the Misallocation of Talent: Evidence from the STEM Doctoral Labor Market." Unpublished.
- Dimmock, Stephen G., Jiekun Huang, and Scott J. Weisbenner.** 2022. "Give Me Your Tired, Your Poor, Your High-Skilled Labor: H-1B Lottery Outcomes and Entrepreneurial Success." *Management Science* 68 (9): 6950–70.
- Doms, Mark E., and Ethan G. Lewis.** 2006. "Labor Supply and Personal Computer Adoption." Federal Reserve Bank of Philadelphia Working Paper 06-10.
- Doran, Kirk, Alexander Gelber, and Adam Isen.** 2022. "The Effects of High-Skilled Immigration Policy on Firms: Evidence from Visa Lotteries." *Journal of Political Economy* 130 (10): 2501–33.
- Dustmann, Christian, and Albrecht Glitz.** 2015. "How Do Industries and Firms Respond to Changes in Local Labor Supply?" *Journal of Labor Economics* 33 (3): 711–50.
- Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock.** 2023. "GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models." Unpublished.
- Facchini, Giovanni, Anna Maria Mayda, and Prachi Mishra.** 2011. "Do Interest Groups Affect US Immigration Policy?" *Journal of International Economics* 85 (1): 114–28.
- Fairlie, Robert W., and Magnus Lofstrom.** 2015. "Immigration and Entrepreneurship." In *Handbook of the Economics of International Migration*, Vol. 1, edited by Barry R. Chiswick and Paul W. Miller, 877–911. Oxford, UK: North-Holland.
- Felten, Edward W., Manav Raj, and Robert Seamans.** 2023. "How Will Language Modelers like ChatGPT Affect Occupations and Industries?" Unpublished.
- Fleming, Lee.** 2001. "Recombinant Uncertainty in Technological Search." *Management Science* 47 (1): 117–32.
- Foley, C. Fritz, and William R. Kerr.** 2013. "Ethnic Innovation and U.S. Multinational Firm Activity." *Management Science* 59 (7): 1529–44.
- Freeman, Richard B.** 2015. "Immigration, International Collaboration, and Innovation: Science and Technology Policy in the Global Economy." *Innovation Policy and the Economy* 15: 153–75.
- Fry, Caroline Viola.** 2023. "Bridging the Gap: Evidence from the Return Migration of African Scientists." *Organization Science* 34 (1): 404–32.
- Ganguli, Ina.** 2015. "Immigration and Ideas: What Did Russian Scientists 'Bring' to the United States?" *Journal of Labor Economics* 33 (S1): S257–88.
- Ghosh, Anirban, Anna Maria Mayda, and Francesc Ortega.** 2014. "The Impact of Skilled Foreign Workers on Firms: An Investigation of Publicly Traded U.S. Firms." IZA Discussion Paper 8684.
- Gibson, John, and David McKenzie.** 2011. "Eight Questions about Brain Drain." *Journal of Economic Perspectives* 25 (3): 107–28.
- Glennon, Britta.** 2023. "How Do Restrictions on High-Skilled Immigration Affect Offshoring? Evidence from the H-1B Program." *Management Science*. <https://doi.org/10.1287/mnsc.2023.4715>.
- Glennon, Britta, Francisco Morales, Seth Carnahan, and Exequiel Hernandez.** 2021. "Does Employing Skilled Immigrants Enhance Competitive Performance? Evidence from European Football Clubs." NBER Working Paper 29446.

- Gould, David M.** 1994. "Immigrant Links to the Home Country: Empirical Implications for U.S. Bilateral Trade Flows." *Review of Economics and Statistics* 76 (2): 302–16.
- Head, Keith, and John Ries.** 1998. "Immigration and Trade Creation: Econometric Evidence from Canada." *Canadian Journal of Economics* 31 (1): 47–62.
- Henderson, Rebecca M., and Kim B. Clark.** 1990. "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms." *Administrative Science Quarterly* 35 (1): 9–30.
- Hernandez, Exequiel.** 2014. "Finding a Home away from Home: Effects of Immigrants on Firms' Foreign Location Choice and Performance." *Administrative Science Quarterly* 59 (1): 73–108.
- Hernandez, Exequiel, and Elena Kulchina.** 2020. "Immigrants and Foreign Firm Performance." *Organization Science* 31 (4): 797–820.
- Hira, Ron.** 2018. "Bridge to Permanent Immigration or Temporary Labor? The H-1B Visa Program Is a Source of Both." In *U.S. Engineering in a Global Economy*, edited by Richard B. Freeman and Hal Salzman, 263–83.
- Hunt, Jennifer.** 2011. "Which Immigrants Are Most Innovative and Entrepreneurial? Distinctions by Entry Visa." *Journal of Labor Economics* 29 (3): 417–57.
- Hunt, Jennifer, and Marjolaine Gauthier-Loiselle.** 2010. "How Much Does Immigration Boost Innovation?" *American Economic Journal: Macroeconomics* 2 (2): 31–56.
- Jaffe, Adam B.** 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value." *American Economic Review* 76 (5): 984–1001.
- 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.
- Javorcik, Beata Smarzynska.** 2004. "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages." *American Economic Review* 94 (3): 605–27.
- Javorcik, Beata S., Ça lar Özden, Mariana Spatareanu, and Cristina Neagu.** 2011. "Migrant Networks and Foreign Direct Investment." *Journal of Development Economics* 94: 231–41.
- Kahn, Shulamit, and Megan J. MacGarvie.** 2016. "How Important Is U.S. Location for Research in Science?" *Review of Economics and Statistics* 98 (2): 397–414.
- Keller, Wolfgang.** 2004. "International Technology Diffusion." *Journal of Economic Literature* 42 (3): 752–82.
- Kerr, William R.** 2008. "Ethnic Scientific Communities and International Technology Diffusion." *Review of Economics and Statistics* 90 (3): 518–37.
- Kerr, Sari Pekkala, and William R. Kerr.** 2018. "Global Collaborative Patents." *Economic Journal* 128 (612): F235–72.
- Kerr, Sari Pekkala, and William R. Kerr.** 2020. "Immigrant Entrepreneurship in America: Evidence from the Survey of Business Owners 2007 and 2012." *Research Policy* 49 (3): 103918.
- Kerr, Sari Pekkala, and William R. Kerr.** 2021. "Immigration Policy Levers for US Innovation and Start-Ups." In *Innovation and Public Policy*, edited by Austan Goolsbee and Benjamin F. Jones, 85–116. Chicago: University of Chicago Press.
- Kerr, Sari Pekkala, William R. Kerr, and William F. Lincoln.** 2015. "Firms and the Economics of Skilled Immigration." *Innovation Policy and the Economy* 15: 115–52.
- 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.
- Kerr, William R., William F. Lincoln, and Prachi Mishra.** 2014. "The Dynamics of Firm Lobbying." *American Economic Journal: Economic Policy* 6 (4): 343–79.
- Lee, Saerom, and Britta Glennon.** 2023. "The Effect of Immigration Policy on Founding Location Choice: Evidence from Canada's Start-Up Visa Program." NBER Working Paper 31634.
- Lewis, Ethan.** 2011. "Immigration, Skill Mix, and Capital Skill Complementarity." *Quarterly Journal of Economics* 126 (2): 1029–69.
- Li, Ashlee.** 2021. "High-Skilled Immigrant Workers and U.S. Firms' Access to Foreign Venture Capital." Paper presented at Association for Public Policy Analysis and Management Conference, Austin, TX, March 27.
- MacGarvie, Megan.** 2006. "Do Firms Learn from International Trade?" *Review of Economics and Statistics* 88 (1): 46–60.
- Marshall, Alfred.** 1890. *Principles of Economics*. London: Macmillan and Co.

- Mayda, Anna Maria, Gianluca Orefice, and Gianluca Santoni. 2022. "Skilled Immigration, Task Allocation and the Innovation of Firms." CESifo Working Paper 10076.
- Mayda, Anna Maria, Francesc Ortega, Giovanni Peri, Kevin Shih, and Chad Sparber. 2020. "New Data and Facts on H-1B Workers across Firms." In *The Roles of Immigrants and Foreign Students in US Science, Innovation, and Entrepreneurship*, edited by Ina Ganguli, Shulamit Kahn, and Megan MacGarvie, 99–121. Chicago: University of Chicago Press.
- Mayda, Anna Maria, Francesc Ortega, Giovanni Peri, Kevin Shih, and Chad Sparber. 2023. "Coping with H-1B Shortages: Firm Performance and Mitigation Strategies." *ILR Review* 76 (5): 919–43.
- McAuliffe, Marie, and Binod Khadria, eds. 2020. *World Migration Report 2020*. Geneva: International Organization for Migration.
- Mehra, Mishita, and Hewei Shen. 2022. "Skilled Immigration, Firms, and Policy." *Journal of International Economics* 139: 103622.
- Melitz, Marc J. 2003. "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.
- Mitaritonna, Cristina, Gianluca Orefice, and Giovanni Peri. 2017. "Immigrants and Firms' Outcomes: Evidence from France." *European Economic Review* 96: 62–82.
- Morales, Nicolas. 2019. "High-Skill Migration, Multinational Companies, and the Location of Economic Activity." *Review of Economics and Statistics*. [https://doi.org/10.1162/rest\\_a\\_01327](https://doi.org/10.1162/rest_a_01327).
- Moser, Petra, Alessandra Voena, and Fabian Waldinger. 2014. "German Jewish Émigrés and US Invention." *American Economic Review* 104 (10): 3222–55.
- Olney, William W. 2012. "Offshoring, Immigration, and the Native Wage Distribution." *Canadian Journal of Economics* 45 (3): 830–56.
- Olney, William W., and Dario Pozzoli. 2021. "The Impact of Immigration on Firm-Level Offshoring." *Review of Economics and Statistics* 103 (1): 177–95.
- Ottaviano, Gianmarco I. P., Giovanni Peri, and Greg C. Wright. 2013. "Immigration, Offshoring, and American Jobs." *American Economic Review* 103 (5): 1925–59.
- Page, Scott E. 2011. *Diversity and Complexity*. Princeton, NJ: Princeton University Press.
- Papademetriou, Demetrios G., and Kate Hooper. 2019. *Competing Approaches to Selecting Economic Immigrants: Points-Based vs Demand-Driven Systems*. Washington, DC: Migration Policy Institute.
- Parrotta, Pierpaolo, Dario Pozzoli, and Mariola Pytlikova. 2014. "Labor Diversity and Firm Productivity." *European Economic Review* 66: 144–79.
- 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 Francisco Requena-Silvente. 2010. "The Trade Creation Effect of Immigrants: Evidence from the Remarkable Case of Spain." *Canadian Journal of Economics* 43 (4): 1433–59.
- Peri, Giovanni, and Chad Sparber. 2009. "Task Specialization, Immigration, and Wages." *American Economic Journal: Applied Economics* 1 (3): 135–69.
- Poole, Jennifer P. 2013. "Knowledge Transfers from Multinational to Domestic Firms: Evidence from Worker Mobility." *Review of Economics and Statistics* 95 (2): 393–406.
- Rauch, James E. 2001. "Business and Social Networks in International Trade." *Journal of Economic Literature* 39 (4): 1177–1203.
- Rauch, James E., and Vitor Trindade. 2002. "Ethnic Chinese Networks in International Trade." *Review of Economics and Statistics* 84 (1): 116–30.
- Roach, Michael, and John Skrentny. 2019. "Why Foreign STEM PhDs Are Unlikely to Work for US Technology Startups." *Proceedings of the National Academy of Sciences of the United States of America* 116 (34): 16805–10.
- Saxenian, AnnaLee. 2006. *The New Argonauts: Regional Advantage in a Global Economy*. Cambridge, MA: Harvard University Press.
- Sharma, Rishi, and Chad Sparber. 2020. "Buying Lottery Tickets for Foreign Workers: Search Cost Externalities Induced by H-1b Policy." IZA Discussion Paper 13892.
- Singh, Jasjit. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science* 51 (5): 756–70.
- Singh, Jasjit. 2008. "Distributed R&D, Cross-Regional Knowledge Integration and Quality of Innovative Output." *Research Policy* 37 (1): 77–96.

- Wang, Dan.** 2015. "Activating Cross-border Brokerage: Interorganizational Knowledge Transfer through Skilled Return Migration." *Administrative Science Quarterly* 60 (1): 133–76.
- Waugh, Michael E.** 2018. "Firm Dynamics and Immigration: The Case of High-Skilled Immigration." In *High-Skilled Migration to the United States and Its Economic Consequences*, Gordon H. Hanson, William R. Kerr, and Sarah Turner, 205–38. Chicago: University of Chicago Press.
- Zwetsloot, Remco, James Dunham, Zachary Arnold, and Tina Huang.** 2019. *Keeping Top AI Talent in the United States*. Washington, DC: Center for Security and Emerging Technology.

# Migration and Innovation: Learning from Patent and Inventor Data

Francesco Lissoni and Ernest Miguelez

**M**igration and innovation have gone together since the dawn of human history. It was migration from the Near East that brought farming to the European continent in the Neolithic period (Skoglund et al. 2012). About 10,000 years later, in 1685, it was Huguenot refugees escaping persecution in France after the revocation of the Edict of Nantes that brought to Prussia the most advanced textile technologies (Scoville 1952; Hornung 2014). In the US experience, a well-known episode is that of the scientists of Jewish origins who, fleeing Germany in the 1930s after the Nazi party's rise to power, brought their knowledge in several fields of chemistry and physics (Moser, Voena, and Waldinger 2014).

But migrants contribute to innovation not only when moving jointly from a more to a less advanced economy, as in these and other well-known historical examples. Many of them travel in the opposite direction and on an individual basis, in search of the best place to develop their own new ideas. The case of Katalin Karikó, who

■ *Francesco Lissoni is Professor of Economics at the Bordeaux School of Economics, University of Bordeaux, France. He is also a Fellow of ICRIOS-Università Bocconi and of the Department of Economic Policy of the Università Cattolica del Sacro Cuore, both in Milan, Italy. Ernest Miguelez is Research Fellow at the Institute of Public Goods and Policies, CSIC (Consejo Superior de Investigaciones Científicas), Madrid, Spain. He is also a Fellow at the Bordeaux School of Economics, University of Bordeaux, France and the AQR-IREA (Grup Anàlisi Quantitativa Regional—Institut de Recerca en Economia Aplicada), University of Barcelona, Barcelona, Spain. Both authors are Fellows at the French Collaborative Institute on Migration (Institut Convergences Migrations), Paris, France. Miguelez is the corresponding author at [ernest.miguelez@csic.es](mailto:ernest.miguelez@csic.es).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.1.27>.

received the Nobel Prize in Medicine in 2023, is exemplary. After earning her PhD in Biochemistry in Hungary, she moved to the United States in 1989 for a postdoc and started studying the potential applications of mRNA technologies to vaccines. Following a long spell at the University of Pennsylvania, and many difficulties in getting tenure and funding, in 2013 she joined the German start-up BioNTech, where she contributed, as vice-president, to the development of its Covid-19 vaccine (Kollewe 2020; Johnson 2021).

While Karikó left Hungary after completing her PhD studies, many students worldwide move abroad in order to earn an advanced degree, especially in scientific fields or engineering. They take advantage of lower transport costs, the spread of English as both an instruction and working language, and the global search for talent by multinational companies and universities (Kerr 2018). Unlike the experienced Huguenot weavers of the seventeenth century or the German scientists in the twentieth, when they first move they do not bring along any superior knowledge, but like Karikó, may be deciding to pursue a research career and swell the ranks of scientists and inventors in the host countries. In the United States, immigrants are overrepresented relative to natives not only among the most productive researchers and the winners of important scientific prizes (Stephan and Levin 2001; Bernstein et al. 2022), but more generally among the students who, upon graduation, end up filing a patent (Hunt and Gauthier-Loiselle 2010) or otherwise hold jobs in science and technology (Hanson and Slaughter 2017). Absent them, R&D-intensive economies would face serious job shortages. In a small, highly innovative country like Switzerland, more than a quarter of scientists and inventors are foreign-born (Cristelli and Lissoni 2020). Absent the possibility of migrating, many of them would have not undertaken a research career or would have not been as productive as they are.

Technology transfer and self-selection (into science and technology jobs and especially those best matching one's own skills) are thus key factors behind migrants' contribution to innovation. An additional factor is the diversity that migrants bring about. While it is difficult to establish a clear-cut causality link, the correlation between migration-induced diversity and innovation is well-documented at the regional or city levels (Bosetti, Cattaneo, and Verdolini 2015) and, in part, also for firms and teams (Parrotta, Pozzoli, and Pytlikova 2014). Anecdotes are uncountable. While Karikó's Nobel prize cowinner and colleague at the University of Pennsylvania, Drew Weissman, was born and bred in the United States, the BioNTech co-founders who reached out to her from Germany were Uğur Şahin and Özlem Türeci, both from Turkish parents. The CEO of Pfizer, BioNTech's key US partner, is Albert Bourla, born in Thessaloniki (Greece), where he first joined the company after earning a PhD in veterinary science. The CEO of Moderna, another US company with a pioneer Covid-19 vaccine, is Stéphane Blancel, born in Marseille (France), with a master's degree in engineering earned in Paris and a subsequent degree at the University of Minnesota.

All three channels through which migration can affect innovation are the objects of an increasing number of empirical studies, many of which make use

of patent and inventor data. Patent data have limitations: not all inventions are patented and many inventions protected by patents turn out not to be worth developing or commercializing. Yet they are a very detailed source of information on both technologies and the individuals behind them, especially in R&D-intensive economies.

In this paper, we first discuss how to access and treat patent data to extract information on inventors in general and migrant ones in particular. We also produce some descriptive statistics that both summarize important facts and serve as cautionary reminders of the data limitations. We emphasize that migrant inventors are a heterogeneous group, which includes not only senior figures holding precious knowledge assets and capable of transferring them across countries, but also junior figures, who may have little knowledge to transfer upon arrival in the host country yet may play a key role in easing R&D labor shortages, swelling the ranks of the highly productive researchers and increasing diversity at the team, firm, and city levels.

We then discuss the empirical research on both roles. When discussing the available evidence on migrant inventors' role in international knowledge diffusion, we emphasize how these research methods come from the convergence of different traditions, such as economic geography and migration economics, whose contact point is the emphasis they place on interpersonal exchanges as key means of knowledge diffusion.

## **Migrant Inventors: Data and Facts**

Quantifying migrant inventors and their contribution to innovation is not a straightforward task, first and foremost because “inventor” is not an occupation listed in professional registers or official national statistics, let alone international ones. However, a functional definition of inventor would refer to those named as such on patent documents, which in most countries must indicate both the generalities of the assignee (the legal owner of the exclusivity rights claimed in the patent) and those of the physical person who conceived the new technology described in the application (the inventor, most often more than one per patent). Assignees and inventors may coincide, but in most cases they do not, with the former being the firm or other organization employing the latter in its R&D labs. Besides, in most countries (including the United States since 2012), it is the assignee that files the patent application and is referred to, in legal jargon, as “applicant.” This convention has implications that we discuss below.

The inventor information reported in patents (when available) includes family and given names as well as the country and place of residence. The same is true for assignees, although for large firms the address may not be that of the R&D laboratory in which the inventors work, but that of the headquarters or a subsidiary dedicated to the management of the company's intellectual property rights. With one exception (discussed later), patents do not report information

on the inventors' birthplace or nationality, from which one could immediately deduce the inventor's status as migrant or native. Before delving into the solutions to this problem, it is useful to recall a few more features of patent data.

### **Patent Data**

With the very recent exception of the Unitary Patent, valid throughout the European Union since June 2023, patents offer a protection of intellectual property rights at the national level, with each country having its own patent office. Applicants seeking protection for the same invention in more than one country need to file for patents in each one of them, which generates what in legal jargon are called "patent families" (Martínez 2010). Such "families" consist of sets of patent documents referring to the same invention and applicant, produced by different patent offices, in different languages and with very similar, but not identical, information (for example, information on inventors may be collected more accurately by some national patent offices than others).

Besides including a very fine technological classification of the inventions (as well as the full text and drawings describing them), patents include cross-references in the form of citations to other patents (Jaffe and de Rassenfosse 2017). Most cross-references have a legal function, as they help the patent examiners to establish how much of the invention for which the patent is requested is new relative to the existing body of scientific and technical knowledge. Other cross-references just serve the purpose of better explaining how the invention works. Many citations are inserted in the patents by the inventors, while others are inserted by the patent attorneys or the examiners themselves. Broadly speaking, economists exploit citations in two main ways. First, they look at the number of citations a patent receives ("backward" citations) as a measure of its importance for follow-up technical advancements. Second, they look at the citations a patent makes to other patents ("forward" citations) as an indirect (and noisy) measure of knowledge diffusion, from the inventors of the cited patents to those of the citing one. These can be combined with measures on the similarity between patents, based on their coclassification in the same technological classes or text analysis (Arts, Cassiman, and Gomez 2018). Patents can also report citations to scientific publications and other documents, which researchers use as measures of knowledge diffusion, in this case from science to technology (Marx and Fuegi 2022).

Because patent offices in countries with meaningful R&D activities (for example, the United States, European countries, or Japan) systematically collect and exchange their data, the geographical coverage of information on inventors is virtually global, with multiple data sources. In what follows, we focus on two of them, namely the US Patent & Trademark Office (USPTO) and the World Intellectual Property Organization (WIPO), a UN agency in charge of easing the application and updating of the international treaties on intellectual property rights.

The US Patent & Trademark Office is today the second-most important office worldwide in terms of yearly patent filings, just behind the Chinese office, and the



most important in cumulative terms. This is due both to the innovation propensity of US firms and to the many foreign companies that seek protection for their intellectual property rights in the United States, in order to export or invest there (WIPO 2022). Many patent applications filed at the US Patent & Trademark Office belong to large patent families—that is, they concern inventions for which patent applications are also filed in other countries. This makes them quite representative of inventive activities worldwide, despite a clear overrepresentation of US-based inventions (de Rassenfosse et al. 2013). All patents granted by the US Patent & Trademark Office can be easily retrieved online from the PatentsView website, produced by the Office of the Chief Economist. PatentsView provides a unique identification for all the inventors listed on such patents, an important piece of information to which we return below.<sup>1</sup>

As for the World Intellectual Property Organization, it collects and publishes information on patent applications that go through the Patent Cooperation Treaty. This treaty, now signed by over 150 countries, allows patent-seekers from signatory countries to file applications in one another’s patent offices by following, at an additional cost, a special procedure. Under some conditions, this treaty is more favorable than that established by older international conventions (WIPO 2023). Generally speaking, patent applications under the Patent Cooperation Treaty are highly relevant for economic research purposes, because they are all meant for international extension and hence generally related to more valuable inventions than those filed only at national offices. In addition, they are treated according to more similar administrative rules, in particular for what concerns their submission procedure. Last but not least, for a bureaucratic accident, the patent applications filed through the procedure of Patent Cooperation Treaty and seeking protection in the United States include, from around the 1980s to 2011, the nationality of the inventors (Migueluez and Fink 2017). These data total a little more than two million patents, but cover 15 percent of all patent families worldwide in the mentioned period. While now outdated, this information continues to be important for both historical and methodological reasons.<sup>2</sup>

<sup>1</sup>Notice that, when referring to patent statistics in general, we speak of “patent applications,” as the available data generally include both those that end up with the patent being granted and those for which the application is rejected by the patent office or abandoned by the applicant. Instead, when referring specifically to data from the US Patent & Trademark Office, we speak of granted patents only. This is because, until the Patent Reform Act of 2005, the US Patent & Trademark Office published only the granted patents and not, like the patent offices of all other countries worldwide, all the applications subject to examination. As a consequence, for patent statistics based on US Patent & Trademark Office data and spanning across year 2005, the best practice is to use only granted patents.

<sup>2</sup>In brief, the Patent Cooperation Treaty procedure requires all applicants to indicate their residence and nationality, at least one of which must correspond to a member state of the treaty (either residence or nationality). At the same time, until the Leahy–Smith America Invents Act of 2011, US law treated all inventors as joint applicants along with the assignee. Hence, on all Patent Cooperation Treaty applications extended to the United States, inventors were required to disclose their nationality, too. When, starting 2012, the US patent system finally introduced a clear distinction between applicants and inventors, assignees took up the role of applicants and the information on the inventors’ nationality disappeared.

### **Defining Inventors**

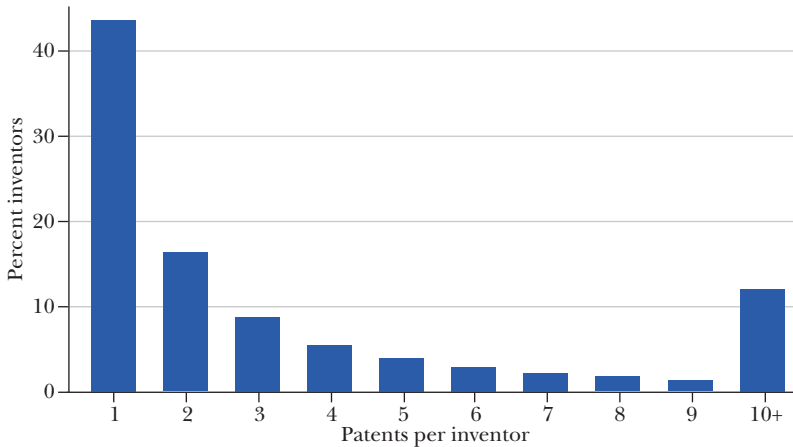
The personal information on inventors that we can extract from patent data is rather noisy. To make it useful for research purposes, it requires two types of treatment: “disambiguation” and “geolocalization.”

Disambiguation consists in assigning a unique identification to each inventor, in order to measure their productivity (number of patents signed) and track their mobility across employers and locations. One specific issue for identifying migrant inventors is that the accuracy of the exercise varies inevitably by linguistic group. In some cases, the difficulty arises from very small numbers of extremely frequent names and surnames, as is the case with East Asian and Scandinavian names. In others, such as with English and Spanish names, identification issues can arise from their diffusion over many countries or continents due to colonial heritage. The available open access data vary across time and by patent office, and the algorithms for disambiguation vary according to the methodology they follow, which in turn affects the incidence of false positives (homonym inventors treated wrongly as the same person) and false negatives (the same inventor treated as two or more different ones). Many disambiguation projects exist, each based on a different set of algorithms and for patents from different offices and time intervals. Only one of them, that of Monath, Jones, and Madhavan (2021) for the US Patent & Trademark Office, has led to a database—PatentsView—that is both freely accessible online and frequently updated.

Geolocalization concerns the relatively simple task of assigning inventors to the residence country whose code is reported by the patent document, as well as more complex operations such as placing them in subnational administrative units (regions, cities, neighborhood or even coordinate points) based on text analysis of the address (de Rassenfosse, Kozak, and Seliger 2019; Maraut et al. 2008; Miguelez et al. 2019; Morrison, Riccaboni, and Pammolli 2017). Geolocalizing inventors is necessary not only to track their mobility in space, but also to second-guess the location of the actual R&D labs in which they have produced their inventions. As explained above, the patent assignee’s address is not always useful to this end, as it may refer to a company’s unit far away from the R&D facilities. The inventors’ residence, instead, can be presumed to be close to their workplace. Hence, the inventor addresses can be used to locate the R&D lab where the invention was produced. For example, when many inventor addresses for patents with the same assignee concentrate in or around the same city, we can presume the R&D lab to be there even if the assignee’s address tells otherwise (for a discussion, see Cristelli and Lissoni 2020).

The results of disambiguation and geolocalization reveal several patterns. For many inventors, patenting is an occasional or intermittent activity. Figure 1 illustrates the distribution of the number of patents per inventor, derived from disambiguated data in the PatentsView database. Approximately half of inventors generate no more than one patent throughout their entire careers, while an additional one-third produce no more than four. The distribution exhibits significant skewness, with an average of approximately 5.3 patents per inventor, a median of 2, and a mode of 1. Some exceptional inventors appear to have signed over 6,000 patents.

Figure 1

**Share of Inventors Grouped by Patents Produced, USPTO, 1976–2020**

*Source:* Authors' own computation based on disambiguated inventor data from Patentsview (US Patent & Trademark Office. "Data Download Tables." PatentsView: <https://patentsview.org/download/data-download-tables>).

*Note:* Figure illustrates the distribution of inventors according to the patents they have produced in their whole inventive lives, derived from disambiguated data in the PatentsView database.

This observation suggests that many inventors are scientists or engineers engaged in industrial R&D for a brief period before transitioning to other professions, while others are practitioners who occasionally come up with inventions based on their experiences on the factory floor. Additionally, there are academics who dabble in technology transfer from time to time. Even the few inventors who consistently file patents over an extended period of time do not do so with a frequency comparable to that of academic scientists who regularly publish scientific research.

These data leave many time gaps in the information provided exclusively by patent documents. In particular, we do not know the inventors' whereabouts and activities before their first patent, after their last one, or between subsequent ones. We can fill these gaps only with survey data or by matching inventor data to a variety of archival sources, such as social security or historical census data. Survey data suggest that most inventors have received an education in science or technology, often at a graduate level. Out of 9,017 European inventors surveyed by Giuri et al. (2007), 77 percent had a university degree and 26 percent a PhD. Toivanen and Väänänen (2012) for Finland and Jung and Ejermo (2014) for Sweden find that, respectively, 67 percent and 76 percent of inventors held a tertiary degree (with 14 percent and 29 percent holding a doctoral degree). The same figures for Japan and the United States, as per Walsh and Nagaoka (2009), are, respectively, 88 percent and 94 percent (13 percent and 45 percent for doctoral degrees). Also, according to Giuri et al. (2007), the most educated inventors are usually found on Chemical and Life Science patents (92 percent with a university degree and 59 percent with a

PhD), while the least educated are in Mechanical Engineering (66.3 percent with a university degree and 9.3 percent with a PhD).

Archival sources suggest that inventors are neither very young, possibly due to the long years many of them spend in education, nor very old, possibly because they are more likely to work in R&D early on in their careers and then move on to other tasks. For the United States, Jones (2009) estimates the average inventor to be aged 30–32 at the time of their first patent, a value confirmed by Cristelli and Lissoni (2020) for foreign inventors in Switzerland. When all patents are considered, the average ages range from 37 in Finland to 46 in the United States (Toivanen and Väänänen 2012; Bell et al. 2019; Akcigit and Goldschlag 2023).

The majority of inventors are men, with women being even more under-represented than in scientific research and, more generally, among science and technology graduates. In 2015, only around 12 percent of all US-based inventors were women (Toole et al. 2019). Causes for this bias are multiple, ranging from women’s weak bargaining position in R&D teams to the influence of role models on their career decisions (Bell et al. 2019).

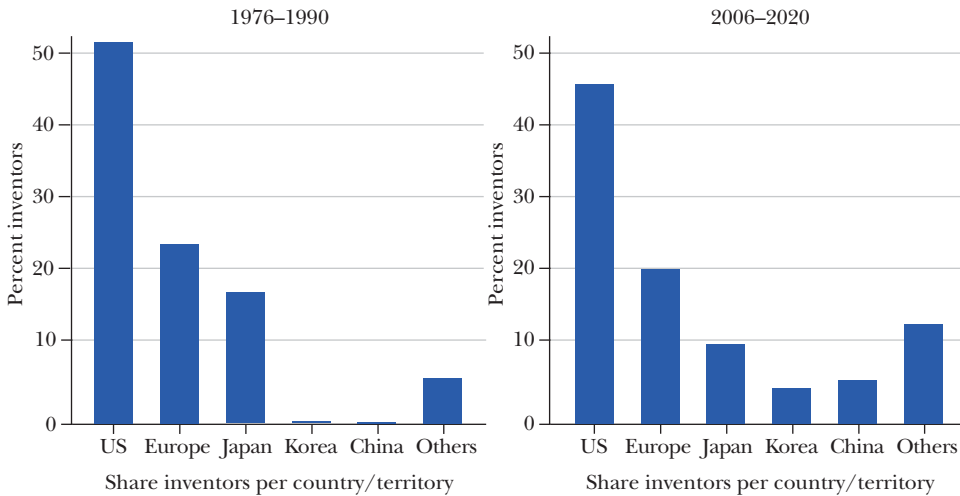
Finally, because patenting is a very uneven geographical phenomenon, both across countries and within them, most inventors are concentrated in a few locations. Figure 2 shows the percent distribution of inventors on US Patent & Trademark Office patents, according to their country of residence, but a similar pattern is found regardless the dataset used (for example, it can be observed from the worldwide analysis using all international activity recently made available by WIPO 2019). Inventors concentrate in a handful of countries such as the United States, European countries, Japan, China, and South Korea, as well as in a few cities and regions therein (Miguelez et al. 2019). In the last 15 years or so, the share of Korea, China, and other countries has been rising, while the share of the United States and Japan has been falling.

This broad portrait of inventors has important implications. First, the supply of inventors is likely to be rigid in the short- and even the medium-term. Becoming an inventor requires, at least in some fields, a high and increasing number of years of scientific or technical education, which cannot be adjusted rapidly and are not accessible to everyone. It may also require pushing back against a diffused gender bias. Second, the demand for inventors is disproportionately concentrated in space, which suggests the importance of a high degree of internal and international mobility of the highly educated. Third, inventors are likely to be in short supply because only a few of them keep patenting throughout their working lives. At the same time, among those who do it, we find the technological leaders who may play an important role in knowledge diffusion.

### **Defining Migrant Inventors**

Assigning migrant status to inventors involves two main difficulties. First, there is the conceptual problem of whether migrants should be defined as such according to their nationality, country of birth, or education place. The appropriate answer will depend both on data availability and the nature of the migration-innovation

Figure 2

**Distribution of Inventors' Country of Residence, USPTO Data**

*Source:* Authors' own computation based on geolocalized inventor data from Patentsview (US Patent and Trademark Office. "Data Download Tables." PatentsView: <https://patentsview.org/download/data-download-tables>). "Europe" refers to the sum of the member states of the European Patent Convention: <https://www.epo.org/en/about-us/foundation/member-states>.

*Note:* Figure shows the percent distribution of inventors on US Patent & Trademark Office patents, according to their country/territory of residence, in two five-year different time windows.

connection we want to study. For example, a country-of-birth definition for migrant inventors may include individuals who moved to the host country at an early age and were entirely educated there, which for some research questions makes them part of the local supply of R&D labor rather than the foreign one. For other research questions, it might make sense to include second-generation migrants or even members of longstanding diasporas (to whom we will refer as "foreign-origin" inventors), as long as they maintain strong ties with their countries of origin and we are interested in the extent to which they send "knowledge remittances" back to origin countries.

Second, practical difficulties arise in retrieving the necessary information from patent documents. As already mentioned, we can find the inventors' nationality only on applications that went through the Patent Cooperation Treaty procedure, up until 2011. Besides time truncation, information on nationality is limited to a specific time period, given by the patent filing dates. As a result, we lack details about the length of the migrant inventor's stay in the host country both before and after the filing date, as well as the type of visa the inventor may have held upon entry. Moreover, nationality can be acquired, meaning that many inventors in these data who are treated as natives may, in fact, be naturalized migrants. This is especially true for those with numerous patents over an extended period, as they are more likely to have resided in the host country long enough to gain citizenship.

Without access to nationality information, we can resort to three alternative strategies for identifying migrant inventors: (1) examining the cross-country mobility of inventors with multiple patents; (2) conducting name analysis to infer the ethnic background of inventors; (3) cross-referencing inventor data with archival sources containing relevant biographical information. We discuss these in turn.

The first approach is an extension of early work on inventors' mobility across firms and organizations to the analysis of mobility first across cities or clusters, then across countries too. It has been used, among others, by Prato (2022) and Bahar et al. (2022, 2023). For an inventor to be qualified as an international migrant, that person must appear on at least two patents with as many or more addresses in different countries.

The second approach, name analysis, has been pioneered by Kerr (2008). It infers the inventors' likely country or region of origin from names and surnames, based on extensive data libraries on their frequency across and within countries. For example, Breschi, Lissoni, and Miguelez (2017) and Coda-Zabetta et al. (2021) make use of IBM's Global Name Recognition system (IBM-GNR), which associates each available name and surname to a vector of countries in which that name or surname is attested, plus information on its frequency inside each country and across all of them. Name analysis may be also used to complement the inventor's mobility method, in order to distinguish between immigrants (those moving to a country where their name and surname are not frequent) from returnees (whose name and surname are frequent), as in Bahar et al. (2023). Similarly, it may help detecting migrants who acquired local nationality on patents that went through the Patent Cooperation Treaty process (Ferrucci and Lissoni 2019).

A third and more recent approach involves linking patent and inventor data with administrative records in the host country, such as Social Security datasets, tax records, censuses, or extensive employee surveys. Bernstein et al. (2022) match US-based inventors with patents from the US Patent & Trademark Office with a commercial database that provides information on over 230 million adult US residents, including their year of birth and Social Security number, the first few digits of which indicate when the number was obtained. Because most US-born individuals receive their Social Security number at birth, while immigrants do so upon their arrival, the authors assume that all the matched inventors whose year of birth differs from that in their Social Security records can be treated as immigrants (whose home country, however, they cannot identify). In particular, they focus on those who obtained their Social Security number after their twentieth birthday. In a different approach, Akcigit and Goldschlag (2023) link disambiguated US-based inventors from the Patentsview database to person-level identifiers, such as the Protected Identification Keys (PIKs) produced by the US Census Bureau. Then, they can match inventors to other data sources, such as the Decennial Census, the American Community Survey, or the Longitudinal Employer Household Dynamics, from which one can obtain information on the inventors' country or region of birth. Examples of this approach for other countries include Toivanen and Väänänen (2012) for

Finland, Jung and Ejermo (2014) for Sweden, and Cristelli and Lissoni (2020) for Switzerland.

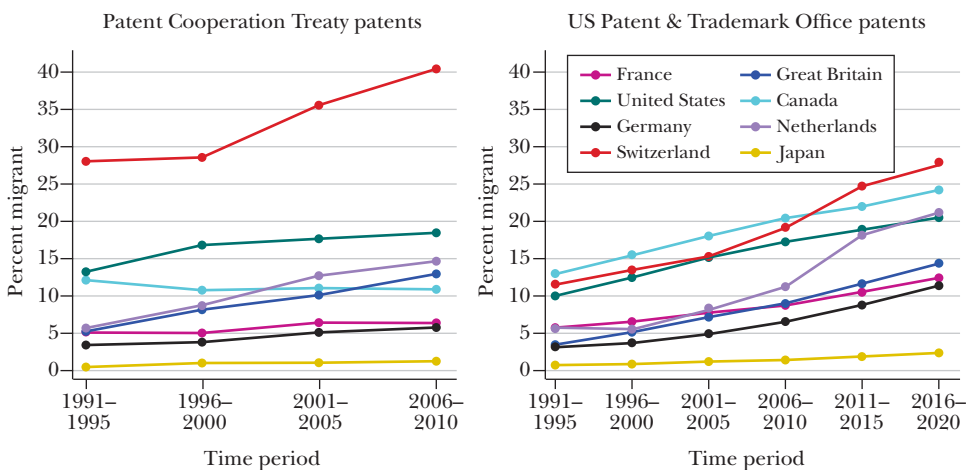
These different approaches all have pros and cons, and the choice between them may depend on the research question at hand.

Computing migration through inventor mobility is straightforward, as one only needs the data extracted from patent applications. Therefore, it can be used for many technologies and countries and over long time periods. Yet, besides requiring careful disambiguation, this method results in oversampling the highly productive and senior migrant inventors relative to the less productive and junior ones. Ethnic classification of inventors based on name analysis is more helpful in this respect, but only under certain circumstances. It misses out on migrants who move between countries with the same dominant language—for example Great Britain and the United States or Germany and Switzerland, just to name two very important migration corridors. In this case, the method necessarily underestimates the number of foreign inventors migrating from one country to the other. At the same time, in other cases, it may lead to an overestimation, as in country pairs with a long migration history that has left noticeable traces in the current distribution of surnames in the destination country (like Italian, German, or Scandinavian names in the United States). This situation bears the risk of blending migrant inventors and native inventors with distant foreign ancestry. While cross-referencing with first names can aid in mitigating this error, as they are better indicators of ethnic identity, it cannot completely eliminate the potential for misclassification.

Linking patent data with administrative records provides the most accurate information, but it also has some drawbacks. First, it can be costly and time-consuming. This applies not only to the creation of new datasets, but also to the updating of existing ones. Second, access to detailed administrative data is possible only for a limited number of host countries. Together, these drawbacks pose obstacles for replicability over time and across countries. In addition, not all the archival resources report the migrants' country of origin nor the age at which they migrated, and thus do not reveal where the inventor was educated.

In conclusion, each of these the approaches might be valid under certain circumstances. Users need to assess what type of error a given approach might produce relative to the research question asked. For example, studying the effect of migrant inventors on the diversity and productivity of inventor teams would require administrative data reporting the inventors' country of birth or education. In the absence of such data, name analysis may provide a second-best approach, but only for host countries that receive most of their migrants from countries with a different language, and the inventor mobility approach would be useless, as most teams include a majority of one-time inventors with no mobility record. However, inventor mobility may be of some use when the research focuses on more senior inventors and their role in knowledge diffusion.

Figure 3

**Presence of Migrant Inventors in Selected High-Intensity R&D Countries, 1991–2020**

Source: Authors' own computation based on geolocalized inventor data from the World Intellectual Property Organization (Miguelez and Fink, 2017) (left panel) and from Patentsview (US Patent and Trademark Office, "Data Download Tables," <https://patentsview.org/download/datadownload-tables>) (right panel).

Note: Figure shows the percent of inventors that are of foreign origin, for a selection of eight high-intensity R&D countries. The figure uses data from the Patent Cooperation Treaty in the left-hand side panel (1991–2010) and data from the US Patent and Trademark Office in the right-hand panel (1991–2020), by five-year intervals.

**How Many Migrant Inventors? And Where?**

We can compare the outcome of different methods for counting migrant inventors. In Figure 3, for example, we combine different patent data sources and migrant detection methods, namely nationality as reported on Patent Cooperation Treaty data (left-hand panel) and name analysis (right-hand panel). Time units are five-year intervals ranging from 1991 to 2010 in the left-hand panel (due to data constraints) and up to 2020 in the right-hand panel.

Whatever method we use, we find that the weight of migrant inventors has been increasing over the entire period in North America and Europe, while remaining negligible in East Asia. But we also observe some differences that may be indicative of measurement error in one or another method. This is evident in the case of Switzerland, which primarily receives immigrants from neighboring countries (France, Germany, and Italy), with which it shares its official languages. Moreover, Switzerland grants its nationality to foreigners rather sparingly. Consequently, the figures obtained using name analysis (right-hand panel) are notably lower compared to those based on nationality (left-hand panel), with the name analysis method being the more unreliable in this case. Canada presents a different scenario, where acquiring nationality is more straightforward, and a substantial number of immigrants come from countries that do not predominantly speak



English or French. In this case, the left-hand panel suggests that the share of immigrant inventors has remained constant over time, while the right-hand panel indicates a growing trend, more in line with what we know about highly skilled immigration trends.

For the United States, the nationality-based estimates suggest a higher share of migrant inventors than the name-based ones, but also a slower growth. This estimate is not far from the figures based on administrative records and concerning only adult immigrants (16 percent in 1990–2016 according to Bernstein et al. 2022), but much lower than those including also child immigrants (24 percent in 2000 and 35 percent in 2016 according to Akcigit and Goldschlag 2023).

The detailed information contained in patent data makes it possible to assess whether migrant and native inventors differ in terms of productivity, specialization, and access or contribution to global knowledge. For instance, Bernstein et al. (2022) estimate that migrants represent 16 percent of all inventors in the United States, but produce 23 percent of US patents, due to their superior productivity. Our own estimates also suggest that migrant inventors have a higher than average productivity, but only in the United States. This may be due to the most productive or promising inventors worldwide having an incentive to move to the United States, the United States offering especially good conditions for migrant inventors to thrive, or a combination of both.

In what areas of science and technology are migrant inventors especially important? Figure 4 reports the share of migrant inventors in different technological fields. Across all countries we consider, this share is higher in the fields of Chemistry and Electrical Engineering, which include many science-based technologies such as Basic communication processes, Micro-structural and nano-technology, Semiconductors, Digital communication, Pharmaceuticals, Computer technology, Organic fine chemistry, and Biotechnology. Such patterns suggest that university education in the host country may be a key channel for immigration, as discussed above.

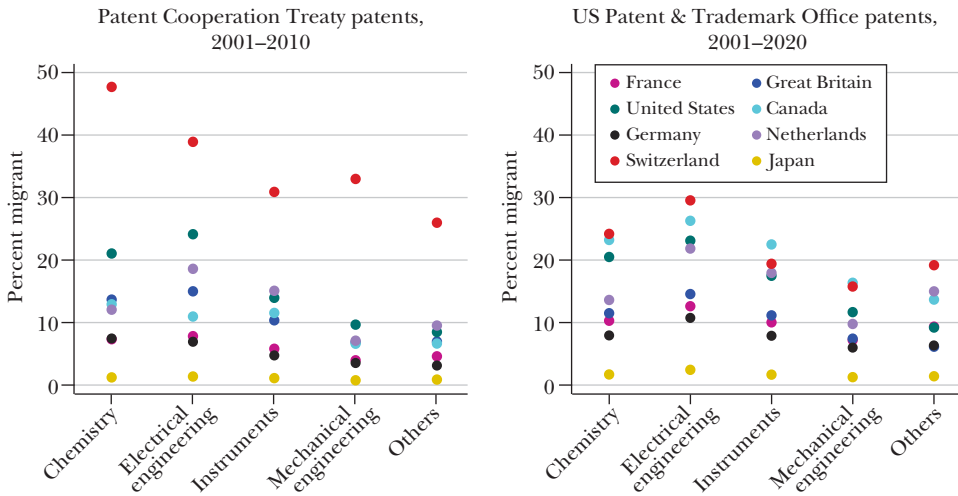
Do migrant inventors play a special role in the international diffusion of knowledge? A cursory look at patent citations suggest that this may be the case. Figure 5 shows the average percentage of forward (left panel) and backward (right panel) foreign citations of US-based inventors, using data only from the US Patent & Trademark Office. As can be seen, in all technological fields, immigrants tend to be systematically more visible internationally (their patents are more cited internationally than those by natives), and tend to fish from a more global pool of knowledge (their patents cite more foreign patents than those by natives). This pattern indicate that migrant inventors belong to a more internationalized community than do native ones.

## **Migrant Inventors as International Knowledge Carriers**

Knowledge diffusion is, by far, the most extensively researched topic in migration and innovation. Its predominance stems from the theoretical importance attached to the study of tacit knowledge as a source of innovation in both economics

Figure 4

**Presence of Migrant Inventors in Selected High-Intensity R&D Countries, 1991–2020, by Broad Technology Field**



Source: Authors' own computation based on geolocalized inventor data from the World Intellectual Property Organization (Miguelez and Fink 2017) (left panel) and from Patentsview (US Patent and Trademark Office. "Data Download Tables." PatentsView: <https://patentsview.org/download/datadownload-tables>) (right panel).

Note: Figure shows the percent of inventors that are of foreign origin, for a selection of eight high-intensity R&D countries. The figure uses data from the Patent Cooperation Treaty in the left-hand side panel (2001–2010) and data from the US Patent and Trademark Office in the right-hand panel (2001–2020), by five broad technology fields.

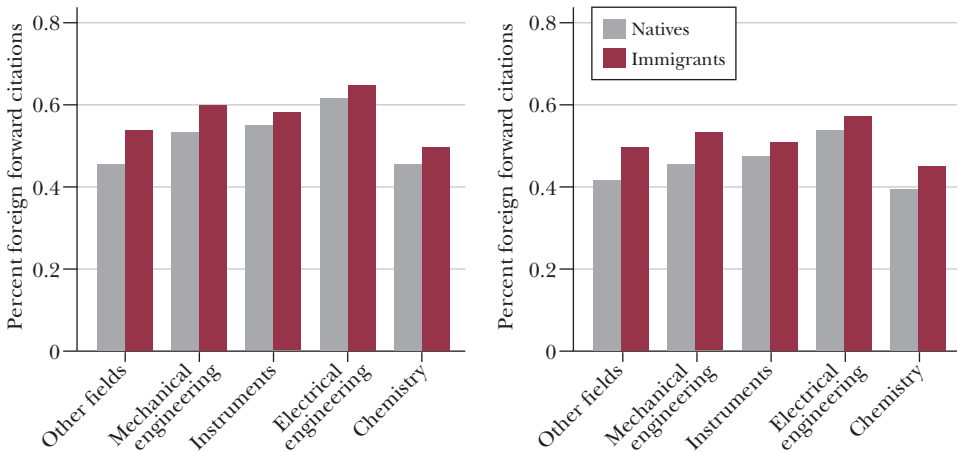
and business studies and to the role played by individuals in disseminating it across organizations and physical locations. The international migration of inventors can be seen as an extreme form of mobility among a highly relevant category of knowledge workers, which lends itself to be studied with migrant inventor data.

### Geographic and Social Proximity

In a classic study using patent citations as knowledge diffusion indicators, Jaffe, Trajtenberg, and Henderson (1993) showed the higher probability for two inventors located in the same metropolitan statistical area within the United States to cite one another's patents, relative to the same probability for two inventors who are not colocated. This finding suggested that knowledge spillovers are concentrated in space and favor the agglomeration of innovative activities in certain locations, due to the crucial role of interpersonal communications. But while personal exchanges are expected to take place more frequently, all else being equal, between individuals closely located in space, it is clear that what matters most is social rather than physical proximity. Two colocated inventors with no contractual or moral obligations to exchange their personal knowledge would not

Figure 5

**Share of US-based, Internationally Cited (Citing) Patents, by Inventors' Migrant Status**



Source: Authors' own computation based on disambiguated inventor and citation data from Patentsview (US Patent & Trademark Office, "Data Download Tables," <https://patentsview.org/download/data-download-tables>).

Note: Figure shows the average percentage of forward (left panel) and backward (right panel) foreign citations of US-based inventors, using data only from the US Patent & Trademark Office. As can be seen, in all technological fields, immigrants tend to be systematically more visible internationally (their patents are more cited internationally than those by natives), and tend to fish from a more global pool of knowledge (their patents cite more foreign patents than those by natives).

do so; for two distant inventors, instead, such obligations may be binding (Breschi and Lissoni 2001). Some indirect evidence supporting this interpretation comes from studies showing that inventors who move apart after a colocation spell still have a higher-than-expected probability to cite each other's subsequent patents (Agrawal, Cockburn, and McHale 2006); as well as from similar findings for former co-inventors or inventors at close distance on a collaboration chain, other things being equal (Breschi and Lissoni 2009).

Subsequent searches for other sources of social proximity with similar effects brought ethnic ties under the spotlight. Such ties may proxy for otherwise unobservable communications between migrant inventors and other co-ethnic inventors, whether located in the same host countries, in different host countries, or in the country of origin. For example, based on a combination of name analysis and the Jaffe, Trajtenberg, and Henderson (1993) methodology, Agrawal, Kapur, and McHale (2008) identify a number of patents by Indian-origin inventors in the United States and show how these are disproportionately cited by other US-based inventors of the same origin, after controlling for their spatial distribution and specialization. Agrawal et al. (2011) extend the analysis to citations from India. The evidence on a role of ethnic ties considerably weakens, but it does not disappear, especially for

patents in Computers & Communications and/or owned by multinational firms, as well as for very important (highly cited) patents.

Similar analysis has been applied to other groups of migrant and foreign-origin inventors in the United States, including Chinese, Koreans, Iranians, Russians, and some Western European countries. Again based on name analysis, Breschi, Lissoni, and Miguelez (2017) find that ethnic ties between migrant inventors matter for Asian and Russian inventors. As for evidence concerning diffusion back to the home countries, this appears significant for China and Russia, but not for India. Instead, patents by Indian-origin inventors in the United States appear to be disproportionately cited by other members of the Indian diaspora, such as those residing in the United Kingdom or other former Commonwealth countries.<sup>3</sup>

One limitation of this kind of study concerns the reliability of patent citations as indicators of knowledge diffusion. Instead of some knowledge transfer from the inventors of the cited patent to those of the citing one, a citation could indicate the existence of some background knowledge embodied in both patents (Thompson and Fox-Kean 2005; Arora, Belenzon, and Lee 2018). In this case, a disproportionate number of citations between inventors from the same ethnic group could be attributed to their common specialization in a narrow set of technologies, rather than to any privileged access to each other's knowledge. While this common specialization might also result from interactions outside the patenting activity, such as shared education or joint professional experiences, the causal link between proximity and diffusion would be, in such a case, less clear.

It is also possible to track knowledge transfers by focusing on the inventors who move and patent across countries and by examining the technical classification of both their patents and of follow-up ones. Bahar et al. (2023) consider the cumulative patenting activity of around 200 countries in 600 technological fields, from 1975 to 2015. In particular, the authors focus on the first years in which each country starts patenting in any given field. Using disambiguated data from US patents, they track the inventors with foreign experience active in any given country, field, and year and find them to be disproportionately represented in the early years of a country's patenting activity in any given field. Interestingly, the effect is stronger for returnees than for immigrants (with name analysis used to distinguish between the former and the latter).

### **Country-level Data: The Preference and Information Channels**

The migration and innovation studies bear a relationship with broader studies that investigate how international migration affects trade and foreign direct investment. The latter developed a number of concepts and methods that can be extended

<sup>3</sup>The lack of evidence for Europeans (excepting Russians) may be explained by a cohort effect and/or a composition effect. In the first case, name analysis may mistakenly identify as foreign-origin inventors some individuals with only a remote European ancestry, and few or no co-ethnic ties. In the second case, inventors may be temporary migrants working for multinational companies, whose corporate ties matter more than ethnic ones.

to the former. For example, from the migration and trade literature we learn that the two may be linked through two main channels: a preference channel, where migrants increase the host country's demand for home-country products; and, more importantly for our purposes, an information channel, where migrants reduce transaction costs of international trade operations through social networks, business contacts, and improved reputation for both host and home countries (Rauch 2001). These effects are particularly significant the more distant two countries are, whether in the physical, cultural, or legal space (Parsons and Winters 2014). Studies examining the relationship between migration and foreign direct investment flows also reveal that migrants play a crucial role in providing valuable information on investment opportunities, costs, and business contacts in their home or destination countries (Burchardi, Chaney, and Hassan 2019; Hernandez 2014).

Patent and inventor data can contribute to the evidence on the importance of the information channel. The work by Kerr (2008) has paved the way. Based on name analysis, the author looks at migrant inventors in the United States and at their patents' forward citations from abroad (excluding company self-citations). He groups the citations over around 100,000 cells, each cell consisting of two-plus-two dimensions: the foreign origin of the citing inventor and that of the cited one; and the technology class of the citing and cited patents. He finds that cells with inventors from the same country of origin are more populated than mixed cells, controlling for technology, which he interprets as evidence of knowledge diffusion from the United States to the migrant inventors' home countries.

More recent studies have followed up and provided a more global perspective. Miguelez and Temgoua (2020) exploit patent citations across countries in a "gravity equation" framework, in which the cost of transmitting information increases with distance. The authors incorporate cross-country data of migrant inventors, using the nationality information on data from Patent Cooperation Treaty patents. Based on both instrumental variable and Poisson estimates, they find that knowledge remittances from migrant diasporas to their home countries are significant, but that knowledge transferred by migrant inventors to their host nations is negligible, with the exception of knowledge flows occurring within multinational enterprises.<sup>4</sup>

Other studies have investigated whether migration can help countries, regions or cities to diversify their technological portfolio. For example, Bahar, Choudhury, and Rapoport (2020) relate the technological diversification of countries to the overall inflows and outflows of migrant inventors, based on nationality information from data from patent using the Patent Cooperation Treaty process. They find that immigrant inventors make it more likely that their host countries will start patenting in technological fields in which their home countries are specialized. But they do not

<sup>4</sup>Similar to Kerr (2008) and Miguelez and Temgoua (2020), other studies have looked at patent collaborations across countries (international co-inventorship, or Global Collaborative Patents). This alternative indicator is less controversial than citations, and it might be somehow related to the latter, though encompassing a broader phenomenon, like international team formation (Kerr and Kerr 2018; Miguelez 2018).

find consistent evidence that emigrant inventors help their home countries to diversify into technologies in which the host countries are specialized. Following up this research, Di Iasio and Miguelez (2022) adopt a similar approach, with several novelties. First, they account for the uneven settlement of emigrant inventors in their host countries across the space and the different specialization of the cities and regions therein. Second, they account for the different specialization of the emigrant inventors themselves. In this way, results for emigrant inventors become far more robust, especially when other internal means to reach these new technologies, like a portfolio of related technologies, are absent (see also Miguelez and Morrison 2023).

## **Skill Supply and Diversity**

Migration can affect innovation not only through knowledge diffusion, but also by increasing the supply of scientists and engineers in the host country and by increasing diversity, whether at the team, firm, or local level. Despite having received less attention than knowledge diffusion, these two channels also bear on broader topics: supply analysis is closely connected to studies on how immigration affects the labor market of the host countries; and the effects of diversity on innovation are one key aspect of the analysis of diversity in team and multinational settings.

### **Skill Supply: Quantity and Quality**

Migrants may affect the supply of inventors in the host country in two ways. First, they may alleviate any shortage of potential inventors due to the lack of local skills, relative to the demand. Second, they may raise the average skill level of inventors due to positive self-selection. A scarcity of skilled labor, including of scientists and engineers, is perceived as a major issue in several innovation-oriented economies (for a recent discussion, see Branstetter, Glennon, and Jensen 2019). At the same time, it is a highly charged political topic, which touches upon sensitive subjects such as education and immigration policies.

To the extent that a skill shortage exists and constrains innovation, international students who decide to stay in the host country after graduation may alleviate it (Chellaraj, Maskus, and Mattoo 2008). These students generally exhibit a higher propensity than natives to enroll in science and technology programs, with the difference increasing when moving from undergraduate to graduate studies (OECD 2022). In addition, at least in the United States, foreign students in science, technology, engineering, and mathematics tend to specialize in fields closer to industrial R&D, which results in a higher propensity to become inventors (Hunt and Gauthier-Loiselle 2010; Hunt 2011).

These findings support a hypothesis of complementarity between native and migrant graduate students and more generally highly-skilled workers, with the potential to increase the productivity of both groups (Peri and Sparber 2009). Conversely, they suggest a lower risk that skilled immigrants will displace native-born workers or have a negative effect on wages. Based on this premise, easing the

restrictions for foreign graduates to move from a student to a work visa or for young, foreign-educated workers to get one would affect both the number of foreign inventors and the total patenting activity in their host countries.

One such restriction that has received considerable attention in the research literature concerns the access to H-1B visas in the United States. While formally conceived for allowing US employers to recruit foreign workers in all “specialty occupations” (those requiring, among others, some university education), these visas have been extensively sought after by foreign-born PhDs employed in industrial R&D as a first step toward permanent residency (Roach and Skrentny 2021). At the same time, they have come under political fire due to their disproportionate use by information technology companies, including US subsidiaries of foreign ones, allegedly for replacing US natives and foreign resident computer scientists with younger, lower paid immigrant ones, and with no appreciable positive effect on innovation.

The supply of H-1B visas has been subject to both national quotas from sending nations and relatively frequent variations, with many employers often being unable to meet their foreign recruitment targets. Scholars investigating migrant inventors have exploited these variations in quasi-experimental studies.

Based on instrumental variable analysis, Kerr and Lincoln (2010) compare the patenting activity of cities and firms with a different historical record of dependence on inventors with Indian and Chinese names, and find the more dependent ones to be more affected by the H-1B visa supply shocks. More importantly for our discussion, the authors show that this effect is mostly due to variations in the number of patents by migrant inventors, rather than by native ones. This suggests that the former do not produce major externalities in favor of the latter, as would occur with knowledge diffusion, and yet they may prove necessary to R&D staff and other inventive teams. These results are also consistent with related city- and firm-level evidence, respectively by Peri, Shih, and Sparber (2015) and Kerr, Kerr, and Lincoln (2015), on the positive effects of H-1B visa supply shocks on the employment and wages of native highly skilled workers following the recruitment of highly skilled foreign workers.

Contrary to such results, however, Doran, Gelber, and Isen (2022) find no evidence that US firms having obtained one or more H-1B visas through a special lottery mechanism innovate more than other, unsuccessful participants to the same lotteries. They also find that H-1B visa-holders are recruited to replace current employees, which suggests that they are mere substitutes of natives (that is, they do not come with complementary skills).

In a different context, a natural experiment in Switzerland also allows a look at how shifts in immigration policy can affect invention. The Swiss education system pushes many natives toward vocational training, rather than toward graduate education in science and technology. At the same time, Switzerland has a R&D system much larger than one would expect based on its population, which makes it highly dependent on foreign talent (as shown earlier in Figure 3). Despite this dependence, due to internal political reasons, its immigration policy remained quite restrictive

until 1999, when it signed an Agreement for the Free Movement of Persons (AFMP) with the European Union as part of a broader cooperation package. While full implementation of the treaty took eight years, restrictions were waived immediately for a special category of immigrants, namely the cross-border workers commuting daily from France, Germany, Italy, and Austria to the Swiss cities close to the international frontier. This makes it possible to compare, during the implementation years, the Swiss regions receiving these cross-border workers to other Swiss regions with similar economies, but too far from the frontier for being touched by the liberalization of cross-border worker visas.

Based on this comparison, Beerli et al. (2021) find that the Swiss regions close to the international frontier saw their intake of foreign skilled workers increasing substantially relative to the others. Cristelli and Lissoni (2020) find the same for foreign inventors, most of whom had no patenting experience prior to their entry in the Swiss labor market (but subsequently signed all their patents with Swiss companies). These foreign inventors do not appear to have displaced the native ones, and instead have increased the productivity of those already active before their entry through direct collaborations. In addition, patenting in the migrant inventors' countries of origin did not react negatively of the migration outflow, which suggests that the lifting of the immigration restrictions allowed a large number of foreign graduates with backgrounds in science and technology to become inventors, which they would not have done if moving to Switzerland had not been possible.

Besides easing possible skill shortages, migrant inventors may be positively self-selected. According to the classic model by Borjas (1987) and Borjas and Bratsberg (1996), positive self-selection of migrants from country A to B occurs when the skill premium in the latter is higher than in the former, so that it is the best and brightest who both move in first place and are more likely to stay (not to return to A). Were the original skill distribution in the two countries similar enough, it may turn out that migrants will be, on average, more skilled than natives, possibly due to the role played by a few exceptional individuals.

This pattern may hold for the United States. As mentioned earlier, Bernstein et al. (2022) find that migrant inventors contribute disproportionately to their host country's patent production, both in general and with respect to highly cited and valuable inventions. These results hold after controlling for many potentially confounding factors, such as the higher concentration of migrant inventors in science-based technologies (whose patents may be more cited or more valuable than others) and in highly innovative hubs, such as the Silicon Valley, where they can benefit from positive externalities. In addition, migrant inventors appear to be technological leaders, to the extent that they also increase significantly the productivity of their collaborators. It is important to stress that the migrants considered in this study include a large number of individuals who arrived in the United States in their twenties, possibly as students or young researchers. When examining their professional life cycles, the authors find that their productivity advantage over natives builds up over time and reaches its peak in their senior years. That is, the pattern does not reflect, at least not exclusively, a few superstar scientists or technologists



moving to the host country after establishing their reputation at home, but instead represents gifted individuals who build their careers in the host country. Quite interestingly, however, they appear to retain strong links outside the United States, as measured by their higher propensity, relative to natives, to cite foreign prior art and to collaborate with foreign inventors. This is compatible with the findings on the diffusion literature we reviewed in the previous section.<sup>5</sup>

### **Diversity and Innovation**

The literature on migration-induced diversity and innovation is extensive, especially in economic geography (as an example, see Bosetti, Cattaneo, and Verdolini 2015). Most studies find a positive link between diversity and innovation, with the latter being often measured with patent-based indicators. A general understanding of these results is that the interaction between diverse individuals may generate more new ideas, or more path-breaking ones, due to complementarity of the inventors' information sets, cognitive differences, and heuristics (Lazear 2008).

One major limitation of this literature so far has been the difficulty of disentangling the different levels at which this diversity interaction occurs, whether at the team, firm, or spatial level. The answer to this question has theoretical, managerial, and policy implications (Kemeny 2017). For example, one possibility is that the fruits of interaction internal to teams or firms may be largely appropriated by the same teams or firms. On the other hand, interaction outside the firm would generate knowledge spillovers and a potential agglomeration force (Olfert and Partridge 2011). This externality would be complementary, but logically different, to that generated by the local access to a diverse workforce (Brunow and Blien 2014).

Another issue concerns the costs of diversity, such as difficulties in communication, polarization of opinions, and disparity of treatment (Harrison and Klein 2007). These may depend on both the number of countries of origin within a team, organization, or location and the size of value differences between individuals from such countries.

Inventor data may contribute to elucidating these issues insofar the co-inventors appearing on the same patent can be treated as a team, whose diversity can be measured by assessing the number of migrants and their distribution across different countries of origin. In addition, information on the patent assignees and the inventor addresses allows an opportunity to control for firm- and location-level diversity. Ferrucci and Lissoni (2019) use this approach to produce a descriptive study on a large number of inventor teams in both Europe and the United States. They establish the migrant status and the countries of origin of inventors by making use of the nationality information on Patent Cooperation Treaty patents, corrected by name analysis. They measure diversity at the team, firm, and location level, based

<sup>5</sup>Notice that Bernstein et al. (2022) cannot establish whether the migrant inventors' foreign ties are with their home countries or with fellow migrants to other host countries, as they do not have information on their nationality or country of birth.

on a simple fractionalization index and its variations. They find a positive association between diversity and the quality of patents, as measured by the number of citations received over three years after filing, and this applies across team, firm, and location differences. Based on Esteban and Ray (1994), they also calculate a “polarization index” capturing possible differences across ethnic groups in working practices and find that, at least for Europe, a negative association with patent quality exists, which indicates that the existence of inventor diversity comes at some cost.<sup>6</sup>

The diversity-innovation literature also suffers from major identification problems; specifically, very innovative firms or cities may attract talented workers and innovators from all over the world, which could create an upward bias in the estimated effect of diversity. Some remedies are available for studies at the firm and location level, such as shift-share analysis as applied by Parrotta, Pozzoli, and Pytlíkova (2014), as well as other instrumental variables as in Campo et al. (2022). However, no fully satisfactory solution to this identification problem in the context of inventor teams has been found so far. A primary difficulty is that these teams are not stable entities, composed of a fixed set of inventors working on several inventions in a row. Rather, they are project-oriented assembling of individuals, also including many one-time inventors, who will not appear on subsequent patents.

## Conclusion

Migrant inventors are a heterogeneous set of workers, most likely with qualifications or experience in the fields of science, technology, engineering, and mathematics, who appear on patent documents. Senior and highly experienced migrant inventors may play a key role in transferring knowledge from their home countries to their host ones. Such inventors are likely to belong to the right tail of the productivity distribution of inventors and will often have patents filed in both countries.

However, these “star” inventors represent a minority of today’s migrant inventors, who instead consist primarily of those who migrate as international students and young graduates, as Nobel Prize winner Katalin Karikó did when she moved to the United States three years after completing her PhD in Hungary. Migrants like her move more to acquire knowledge rather than to transfer it, and their early postmigration jobs often involve work as junior staff in R&D teams. They contribute positively to innovation in the host countries, to the extent that they come with or are willing to acquire any skill in short supply in the host countries. Their

<sup>6</sup>The simplest fractionalization index is equal to one minus the Herfindahl index of the inventors’ nationality shares, whether in a team, firm, or city. Variations occur mainly by attaching weights to the different shares so to take into account the cultural or linguistic heterogeneity across ethnic groups. For a discussion, see Nijkamp and Poot (2015). As for the polarization index, this measures how distant the team members’ opinions are with respect to different propositions as measured, for example, by a Likert-scale assessment of agreement/disagreement (that is, a scale with a range of perhaps five or seven numerical choices ranging from full agreement to full disagreement).

presence in teams, firms, or cities may also enhance innovation through diversity, to the extent that it sustains creativity. Like Karikó, many migrant inventors are positively self-selected, as long as they move to host countries where their skills and determination are better rewarded than at home. As such, they may turn out to be particularly productive and establish themselves as technological leaders. So far, however, evidence in this respect has been gathered only for the United States.

Even when moving in their junior years, migrant inventors may contribute to knowledge diffusion in two ways. First, by maintaining some links to their home countries—for example, by collaborating with local inventors—they may in be in a better position than native inventors to source foreign knowledge (Miguelez and Morrison 2023). Second, they can contribute to innovation in their home countries via knowledge remittances, but evidence on this point needs to be strengthened.

Future investigations of the effect of migration on the supply of inventors will require not only innovative uses of patent data, but also more investments in matching inventor data to administrative records where evidence is lacking, especially outside the United States. One relatively untapped resource is the information one can extract from the Electronic Theses and Dissertations database, which has already been used to track foreign PhD careers in science (Kahn and MacGarvie 2016). Some early examples of this approach have been conducted for Sweden and Germany (Zheng and Ejermo 2015; Buenstorf, Heinisch, and Kapa 2022).<sup>7</sup> Similarly, one could experiment with retrieving and matching information from social media profiles. This approach may come at a considerable cost in terms of sampling bias, but it may provide otherwise missing information on return patterns (Breschi, Lissoni, and Miguelez 2020).

■ *The authors thank Massimiliano Coda-Zabetta and Gianluca Tarasconi for their help in data production, and the journal's editors for their encouragement to write this piece and the many useful comments to the first drafts. We acknowledge funding from the French National Research Agency (TKC project – ref.: ANR-17-CE26-0016), the French State's framework of the Investments for the Future" programme IdEx université de Bordeaux / GPR HOPE, and the IPR ALLIES (CNRS-CSIC)". Ernest Miguelez also acknowledges financial support from the Spanish Ministerio de Ciencia e Innovación (PID2020-119096RB-I00/MCIN/AEI).*

<sup>7</sup>See also Delgado and Murray (2023), albeit with an application to gender studies.

## References

- Agrawal, Ajay, Iain Cockburn, and John McHale. 2006. "Gone but Not Forgotten: Knowledge Flows, Labor Mobility, and Enduring Social Relationships." *Journal of Economic Geography* 6 (5): 571–91.
- Agrawal, Ajay, Devesh Kapur, and John McHale. 2008. "How Do Spatial and Social Proximity Influence Knowledge Flows? Evidence from Patent Data." *Journal of Urban Economics* 64 (2): 258–69.
- Agrawal, Ajay, Devesh Kapur, John McHale, and Alexander Oettl. 2011. "Brain Drain or Brain Bank? The Impact of Skilled Emigration on Poor-Country Innovation." *Journal of Urban Economics* 69 (1): 43–55.
- Akcigit, Ufuk, and Nathan Goldschlag. 2023. "Measuring the Characteristics and Employment Dynamics of U.S. Inventors." NBER Working Paper 31086.
- Arora, Ashish, Sharon Belenzon, and Honggi Lee. 2018. "Reversed Citations and the Localization of Knowledge Spillovers." *Journal of Economic Geography* 18 (3): 495–521.
- Arts, Sam, Bruno Cassiman, and Juan Carlos Gomez. 2018. "Text Matching to Measure Patent Similarity." *Strategic Management Journal* 39 (1): 62–84.
- Bahar, Dany, Prithwiraj Choudhury, Ernest Miguelez, and S. Signorelli. 2023. "Global Mobile Inventors." Unpublished.
- Bahar, Dany, Prithwiraj Choudhury, and Hillel Rapoport. 2020. "Migrant Inventors and the Technological Advantage of Nations." *Research Policy* 49 (9): 103947.
- Bahar, Dany, Prithwiraj Choudhury, Sara Signorelli, and James M. Sappanfield. 2022. "Talent Flows and the Geography of Knowledge Production: Causal Evidence from Multinational Firms." Harvard Business School Working Paper 22-047.
- Beerli, Andreas, Jan Ruffner, Michael Siegenthaler, and Giovanni Peri. 2021. "The Abolition of Immigration Restrictions and the Performance of Firms and Workers: Evidence from Switzerland." *American Economic Review* 111 (3): 976–1012.
- Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. 2019. "Who Becomes an Inventor in America? The Importance of Exposure to Innovation." *Quarterly Journal of Economics* 134 (2): 647–713.
- Bernstein, Shai, Rebecca Diamond, Abhisit Jiranaphawiboon, Timothy McQuade, and Beatriz Pousada. 2022. "The Contribution of High-Skilled Immigrants to Innovation in the United States." NBER Working Paper 30797.
- Borjas, George J. 1987. "Self-Selection and the Earnings of Immigrants." *American Economic Review* 77 (4): 531–53.
- Borjas, George J., and Bernt Bratsberg. 1996. "Who Leaves? The Outmigration of the Foreign-Born." *Review of Economics and Statistics* 78 (1): 165–76.
- Bosetti, Valentina, Cristina Cattaneo, and Elena Verdolini. 2015. "Migration of Skilled Workers and Innovation: A European Perspective." *Journal of International Economics* 96 (2): 311–22.
- Branstetter, Lee G., Britta Glennon, and J. Bradford Jensen. 2019. "The IT Revolution and the Globalization of R&D." *Innovation Policy and the Economy* 19: 1–37.
- Breschi, Stefano, and Francesco Lissoni. 2001. "Knowledge Spillovers and Local Innovation Systems: A Critical Survey." *Industrial and Corporate Change* 10 (4): 975–1005.
- 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.
- Breschi, Stefano, Francesco Lissoni, and Ernest Miguelez. 2017. "Foreign-Origin Inventors in the USA: Testing for Diaspora and Brain Gain Effects." *Journal of Economic Geography* 17 (5): 1009–38.
- Breschi, Stefano, Francesco Lissoni, and Ernest Miguelez. 2020. "Return Migrants' Self-Selection: Evidence for Indian Inventors." In *The Role of Immigrants and Foreign Students in Science, Innovation, and Entrepreneurship*, edited by Ina Ganguli, Shulamit Kahn, and Megan MacGarvie, 17–48. Chicago: University of Chicago Press.
- Brunow, Stephan, and Uwe Blien. 2014. "Effects of Cultural Diversity on Individual Establishments." *International Journal of Manpower* 35 (1/2): 166–86.
- Buenstorf, Guido, Dominik P. Heimesch, and Matthias Kapa. 2022. "International Doctoral Graduates as Inventors in the German Innovation System." *Industry and Innovation* 29 (4): 488–510.
- Burchardi, Konrad B., Thomas Chaney, and Tarek A. Hassan. 2019. "Migrants, Ancestors, and Foreign Investments." *Review of Economic Studies* 86 (4): 1448–86.
- Campo, Francesco, Mariapia Mendola, Andrea Morrison, and Gianmarco Ottaviano. 2022. "Talents and

- Cultures: Immigrant Inventors and Ethnic Diversity in the Age of Mass Migration.” *Journal of the European Economic Association* 20 (5): 1971–2012.
- Chellaraj, Gnanaraj, Keith E. Maskus, and Aaditya Mattoo.** 2008. “The Contribution of International Graduate Students to US Innovation.” *Review of International Economics* 16 (3): 444–62.
- Coda-Zabetta, Massimiliano, Christian Chacua, Francesco Lissoni, Ernest Miguelez, J. Raffo, and Deyun Yin.** 2021. “The Missing Link: International Migration in Global Clusters of Innovation.” Bordeaux Working Paper 2021-02.
- Cristelli, Gabriele, and Francesco Lissoni.** 2020. “Free Movement of Inventors: Open-Border Policy and Innovation in Switzerland.” Unpublished.
- de Rassenfosse, Gaétan, H el ene DERNIS, Dominique Guellec, Lucio Picci, and Bruno van Pottelsberghe de la Potterie.** 2013. “The Worldwide Count of Priority Patents: A New Indicator of Inventive Activity.” *Research Policy* 42 (3): 720–37.
- de Rassenfosse, Ga etan, Jan Kozak, and Florian Seliger.** 2019. “Geocoding of Worldwide Patent Data.” *Scientific Data* 6: 260.
- Delgado, Mercedes, and Fiona E. Murray.** 2023. “Faculty as Catalysts for Training New Inventors: Differential Outcomes for Male and Female PhD Students.” *Proceedings of the National Academy of Sciences* 120 (36): e2200684120.
- Di Iasio, Valentina, and Ernest Miguelez.** 2022. “The Ties That Bind and Transform: Knowledge Remittances, Relatedness and the Direction of Technical Change.” *Journal of Economic Geography* 22 (2): 423–48.
- Doran, Kirk, Alexander Gelber, and Adam Isen.** 2022. “The Effects of High-Skilled Immigration Policy on Firms: Evidence from Visa Lotteries.” *Journal of Political Economy* 130 (10): 2501–33.
- Esteban, Joan-Maria, and Debraj Ray.** 1994. “On the Measurement of Polarization.” *Econometrica* 62 (4): 819–51.
- Ferrucci, Edoardo, and Francesco Lissoni.** 2019. “Foreign Inventors in Europe and the United States: Diversity and Patent Quality.” *Research Policy* 48 (9): 103774.
- Giuri, Paola, Myriam Mariani, Stefano Brusoni, Gustavo Crespi, Dominique Francoz, Alfonso Gambardella, Walter Garcia-Fontes et al.** 2007. “Inventors and Invention Processes in Europe: Results from the PatVal-EU Survey.” *Research Policy* 36 (8): 1107–27.
- Hanson, Gordon H., and Matthew J. Slaughter.** 2017. “High-Skilled Immigration and the Rise of STEM Occupations in US Employment.” In *Education, Skills, and Technical Change: Implications for Future US GDP Growth*, edited by Charles R. Hulten and Valerie A. Ramey, 465–94. Chicago: University of Chicago Press.
- Harrison, David A., and Katherine J. Klein.** 2007. “What’s the Difference? Diversity Constructs as Separation, Variety, or Disparity in Organizations.” *Academy of Management Review* 32 (4): 1199–1228.
- Hernandez, Exequiel.** 2014. “Finding a Home Away from Home: Effects of Immigrants on Firms’ Foreign Location Choice and Performance.” *Administrative Science Quarterly* 59 (1): 73–108.
- Hornung, Erik.** 2014. “Immigration and the Diffusion of Technology: The Huguenot Diaspora in Prussia.” *American Economic Review* 104 (1): 84–122.
- Hunt, Jennifer.** 2011. “Which Immigrants Are Most Innovative and Entrepreneurial? Distinctions by Entry Visa.” *Journal of Labor Economics* 29 (3): 417–57.
- Hunt, Jennifer, and Marjolaine Gauthier-Loiselle.** 2010. “How Much Does Immigration Boost Innovation?” *American Economic Journal: Macroeconomics* 2 (2): 31–56.
- Jaffe, Adam B., and Ga etan de Rassenfosse.** 2017. “Patent Citation Data in Social Science Research: Overview and Best Practices.” *Journal of the Association for Information Science and Technology* 68 (6): 1360–74.
- 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.
- Johnson, Carolyn Y.** 2021. “A One-Way Ticket. A Cash-Stuffed Teddy Bear. A Dream Decades in the Making.” *Washington Post*, October 1. <https://www.washingtonpost.com/health/2021/10/01/katalin-kariko-covid-vaccines>.
- Jones, Benjamin F.** 2009. “The Burden of Knowledge and the ‘Death of the Renaissance Man’: Is Innovation Getting Harder?” *Review of Economic Studies* 76 (1): 283–317.
- Jung, Taehyun, and Olof Ejermo.** 2014. “Demographic Patterns and Trends in Patenting: Gender, Age, and Education of Inventors.” *Technological Forecasting and Social Change* 86: 110–24.
- Kahn, Shulamit, and Megan J. MacGarvie.** 2016. “How Important Is U.S. Location for Research in Science?” *Review of Economics and Statistics* 98 (2): 397–414.
- Kemeny, Thomas.** 2017. “Immigrant Diversity and Economic Performance in Cities.” *International*

- Regional Science Review* 40 (2) 164–208.
- Kerr, Sari Pekkala, and William R. Kerr.** 2018. “Global Collaborative Patents.” *Economic Journal* 128 (612): F235–72.
- Kerr, Sari Pekkala, William R. Kerr, and William F. Lincoln.** 2015. “Skilled Immigration and the Employment Structures of US Firms.” *Journal of Labor Economics* 33 (S1): S147–86.
- Kerr, William R.** 2008. “Ethnic Scientific Communities and International Technology Diffusion.” *Review of Economics and Statistics* 90 (3): 518–37.
- Kerr, William R.** 2018. *The Gift of Global Talent: How Migration Shapes Business, Economy, and Society*. Stanford: Stanford University Press.
- 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.
- Kollewe, Julia.** 2020. “Covid Vaccine Technology Pioneer: ‘I Never Doubted It Would Work.’” *The Guardian*, November 21. <https://www.theguardian.com/science/2020/nov/21/covid-vaccine-technology-pioneer-i-never-doubted-it-would-work>.
- Lazear, Edward P.** 2008. “Diversity and Immigration.” In *Issues in the Economics of Immigration*, edited by George J. Borjas, 117–42. Chicago: University of Chicago Press.
- Lissoni, Francesco, and Ernest Miguelez.** 2024. “Replication data for: Migration and Innovation: Learning from Patent and Inventor Data.” American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E195744V1>.
- Maraut, Stéphanne, Hélène Dernis, Colin Webb, Vincenzo Spiezia, and Dominique Guellec.** 2008. “The OECD REGPAT Database: A Presentation.” OECD Science, Technology, and Industry Working Paper 2008/02.
- Martínez, Catalina.** 2010. “Patent Families: When Do Different Definitions Really Matter?” *Scientometrics* 86 (1): 39–63.
- Marx, Matt, and Aaron Fuegi.** 2022. “Reliance on Science by Inventors: Hybrid Extraction of In-Text Patent-to-Article Citations.” *Journal of Economics and Management Strategy* 31 (2): 369–92.
- Miguelez, Ernest.** 2018. “Inventor Diasporas and the Internationalization of Technology.” *World Bank Economic Review* 32 (1): 41–63.
- Miguelez, Ernest, and Carsten Fink, eds.** 2017. *The International Mobility of Talent and Innovation: New Evidence and Policy Implications*. Cambridge, UK: Cambridge University Press.
- Miguelez, Ernest, and Andrea Morrison.** 2023. “Migrant Inventors as Agents of Technological Change.” *Journal of Technology Transfer* 48 (2): 669–92.
- Miguelez, Ernest, Julio D. Raffo, Christian Chacua, Massimiliano Coda-Zabetta, Deyun Yin, Francesco Lissoni, and Gianluca Tarasconi.** 2019. “Tied In: The Global Network of Local Innovation.” World Intellectual Property Organization (WIPO) Economic Research Working Paper 58.
- Miguelez, Ernest, and Claudia Noumedem Temgoua.** 2020. “Inventor Migration and Knowledge Flows: A Two-Way Communication Channel?” *Research Policy* 49 (9): 103914.
- Monath, Nicholas, Christopher Jones, and Sarvo Madhavan.** 2021. “PatentsView: Disambiguating Inventors, Assignees, and Locations.” Technical report, American Institutes for Research, Arlington, VA. [https://s3.amazonaws.com/data.patentsview.org/documents/PatentsView\\_Disambiguation\\_Methods\\_Documentation.pdf](https://s3.amazonaws.com/data.patentsview.org/documents/PatentsView_Disambiguation_Methods_Documentation.pdf).
- Morrison, Greg, Massimo Riccaboni, and Fabio Pammolli.** 2017. “Disambiguation of Patent Inventors and Assignees Using High-Resolution Geolocation Data.” *Scientific Data* 4: 170064.
- Moser, Petra, Alessandra Voena, and Fabian Waldinger.** 2014. “German Jewish Émigrés and US Invention.” *American Economic Review* 104 (10): 3222–55.
- Nijkamp, Peter, and Jacques Poot.** 2015. “Cultural Diversity: A Matter of Measurement.” In *The Economics of Cultural Diversity*, edited by Peter Nijkamp, Jacques Poot, and Jessie Bakens, 17–51. Cheltenham, UK: Edward Elgar Publishing.
- OECD.** 2022. *International Migration Outlook 2022*. Paris: OECD Publishing.
- Olfert, Margaret Rose, and Mark Partridge.** 2011. “Creating the Cultural Community: Ethnic Diversity vs. Agglomeration.” *Spatial Economic Analysis* 6 (1): 25–55.
- Parrotta, Pierpaolo, Dario Pozzoli, and Mariola Pytlikova.** 2014. “The Nexus between Labor Diversity and Firm’s Innovation.” *Journal of Population Economics* 27: 303–64.
- Parsons, Christopher R., and L. Alan Winters.** 2014. “International Migration, Trade and Aid: A Survey.” In *International Handbook on Migration and Economic Development*, edited by Robert E. B. Lucas, 65–112. Cheltenham, UK: Edward Elgar Publishing.
- Peri, Giovanni, Kevin Shih, and Chad Sparber.** 2015. “STEM Workers, H-1B Visas, and Productivity in US

- Cities.” *Journal of Labor Economics* 33 (S1): S225–55.
- Peri, Giovanni, and Chad Sparber.** 2009. “Task Specialization, Immigration, and Wages.” *American Economic Journal: Applied Economics* 1 (3): 135–69.
- Prato, Marta.** 2022. “The Global Race for Talent: Brain Drain, Knowledge Transfer, and Growth.” Unpublished.
- Rauch, James E.** 2001. “Business and Social Networks in International Trade.” *Journal of Economic Literature* 39 (4): 1177–1203.
- Roach, Michael, and John Skrentny.** 2021. “Rethinking Immigration Policies for STEM Doctorates.” *Science* 371 (6527): 350–52.
- Scoville, Warren C.** 1952. “The Huguenots and the Diffusion of Technology. II.” *Journal of Political Economy* 60 (5): 392–411.
- Skoglund, Pontus, Helena Malmström, Maanasa Raghavan, Jan Storå, Per Hall, Eske Willerslev, M. Thomas P. Gilbert, Anders Götherström, and Mattias Jakobsson.** 2012. “Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe.” *Science* 336 (6080): 466–69.
- Stephan, Paula E., and Sharon G. Levin.** 2001. “Exceptional Contributions to US Science by the Foreign-Born and Foreign-Educated.” *Population Research and Policy Review* 20: 59–79.
- Thompson, Peter, and Melanie Fox-Kean.** 2005. “Patent Citations and the Geography of Knowledge Spillovers: A Reassessment.” *American Economic Review* 95 (1): 450–60.
- Toivanen, Otto, and Lotta Väänänen.** 2012. “Returns to Inventors.” *Review of Economics and Statistics* 94 (4): 1173–90.
- Toole, Andrew A., Amanda Myers, Charles A. W. deGrazia, Stefano Breschi, Edoardo Ferrucci, Francesco Lissoni, Ernest Miguelez, Valerio Sterzi, and Gianluca Tarasconi.** 2019. “Progress and Potential: A Profile of Women Inventors on U.S. Patent.” Office of the Chief Economist IP Data Highlights 2.
- Walsh, John P., and Sadao Nagaoka.** 2009. “Who Invents? Evidence from the Japan–U.S. Inventor Survey.” RIETI Discussion Paper 09-E-034.
- WIPO.** 2019. *World Intellectual Property Report 2019—The Geography of Innovation: Local Hotspots, Global Networks*. Geneva: WIPO.
- WIPO.** 2022. *World Intellectual Property Indicators 2022*. Geneva: WIPO.
- WIPO.** 2023. *PCT Yearly Review 2023: The International Patent System*. Geneva: WIPO.
- Zheng, Yannu, and Olof Ejeremo.** 2015. “How Do the Foreign-Born Perform in Inventive Activity? Evidence from Sweden.” *Journal of Population Economics* 28: 659–95.x





# Tax Equity in Low- and Middle-Income Countries

Pierre Bachas ☉ Anders Jensen ☉ Lucie Gadenne

**I**ncome inequality is high in developing countries and has either stagnated or increased over the past 30 years. Recent estimates from Africa show that, at the regional level, the share of pre-tax income of the top 10 percent is close to 55 percent (Chancel et al. 2023); similarly high levels of inequality are found in large developing countries for which data are available, including Brazil at 58 percent, China at 43 percent, India at 57 percent, and Indonesia at 47 percent (World Inequality Database at <https://wid.world/>). These levels are comparable to or higher than those observed in the United States, a developed country commonly described as highly unequal, where the share of pre-tax income of the top 10 percent was 46 percent in 2021.

This paper asks what role taxation can or might play in reducing high levels of inequality in low- and middle-income countries. Are policy instruments like the personal income tax or value-added tax used in similar ways as in high-income countries, and if so, do they have comparable distributional effects? Alternatively, do specific features of the economic structure of developing countries dilute, or

■ *Pierre Bachas is Assistant Professor of Finance at the École Supérieure des Sciences Economiques et Commerciales (ESSEC) Business School, Cergy, France. At the time of writing, he was on leave from the Development Research Group, World Bank, Washington, DC. Anders Jensen is Associate Professor of Public Policy, Kennedy School of Government, Harvard University, Cambridge, Massachusetts. Lucie Gadenne is Senior Lecturer, School of Economics and Finance at Queen Mary, University of London, and an Associate Director at the Institute for Fiscal Studies, London, United Kingdom. Their email addresses are [pbachas@worldbank.org](mailto:pbachas@worldbank.org), [anders\\_jensen@hks.harvard.edu](mailto:anders_jensen@hks.harvard.edu), and [lgadenne@gmail.com](mailto:lgadenne@gmail.com).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.1.55>.

even reverse, the redistributive effects of tax instruments that are commonly considered to be progressive or regressive? If so, how should we rethink tax design over the development path?

## **Hard Times: Challenges of Progressive Taxation in Low- and Middle-Income Countries**

When considering the equity effects of a tax system, economists and policy-makers typically want to know three key features about who pays taxes. First, the *statutory incidence* describes the taxes legally enacted by the government, and thus who might be expected to pay taxes directly. For example, many countries implement a personal income tax with statutory marginal tax rates that increase with income—such that the statutory tax burden is higher for more well-off individuals. Second, the possibility of tax evasion (or avoidance) means that the *de facto incidence* of who actually pays taxes will differ from the statutory incidence. The ability (and willingness) to evade taxes will often differ between individuals. For example, well-off individuals may have access to sophisticated evasion or avoidance techniques; this would lower their de facto tax burden (how much in taxes they actually pay) and, if such techniques are less available to those with lower income, reduce the progressivity of the tax system (because the difference in actual tax burdens between rich and poor will be smaller than the difference in statutory tax burdens). Third, the *economic incidence* of taxes refers to the fact that market prices respond to taxes, potentially shifting the burden of taxation away from the agents who remit taxes to those who trade with them. An obvious example is that while a sales tax or a value-added tax is collected from firms, the burden of such taxes is often (at least partially) passed along to consumers in the form of higher prices.<sup>1</sup> As we shall see below, individuals who differ in income also often differ in their market behavior; this, in turn, provides the channel through which economic incidence has equity impacts.

These three features determine the equity effects of a tax system, as they change the distribution of income “pre-tax” versus “post-tax.” While this dimension of equity is the focus of our article, it is important to note that additional equity impacts will also occur when the government spends the collected taxes on public goods and services—which changes the distribution of income “post-tax, pre-transfer” versus “post-tax, post-transfer.”<sup>2</sup>

<sup>1</sup>The most commonly implemented indirect tax around the world is the value-added tax, which levies a specified tax rate on the value-added created at each step of the production chain (including at the retail stage where products are sold to consumers). There is no value-added tax in the United States, where subnational governments implement general sales taxes levied at the retail stage. Most of the evidence on indirect taxes in this paper is based on studies of the value-added tax, but the main intuitions for the distributional impacts of an indirect tax would carry over to a setting with a retail sales tax (under certain assumptions).

<sup>2</sup>Reviewing this dimension of equity lies beyond the scope of this article. See, for example, the work by Commitment to Equity (Lustig 2022).

The distinction between statutory and economic incidence has a long history in public finance, but the concept of de facto incidence has received less attention. This may be because researchers and policymakers in high-income countries often assume (at least implicitly) that most agents who should pay taxes do so (an optimistic assumption, see Slemrod 2007). However, the assumption that tax evasion can be neglected clearly does not hold in low- and middle-income countries, making the concept of de facto incidence important in these settings. We return to it frequently throughout this article.

In the context of low- and middle-income countries, it is possible to document the statutory incidence of tax policies by using databases compiled by international organizations and researchers (these are discussed further below, but for a comprehensive overview, see Lustig 2022). However, evidence from low- and middle-income countries suggests that de facto incidence differs substantially from statutory incidence for both obvious and subtle reasons. Moreover, with few exceptions, work in low- and middle-income countries has little to say about economic incidence—with most papers dealing with the issue through implicit or explicit assumptions.<sup>3</sup> These assumptions are sometimes necessary to make progress, but they can still be problematic: numerous studies in high-income countries show that statutory and economic incidence often differ substantially (for example, Chetty, Looney, and Kroft 2009).

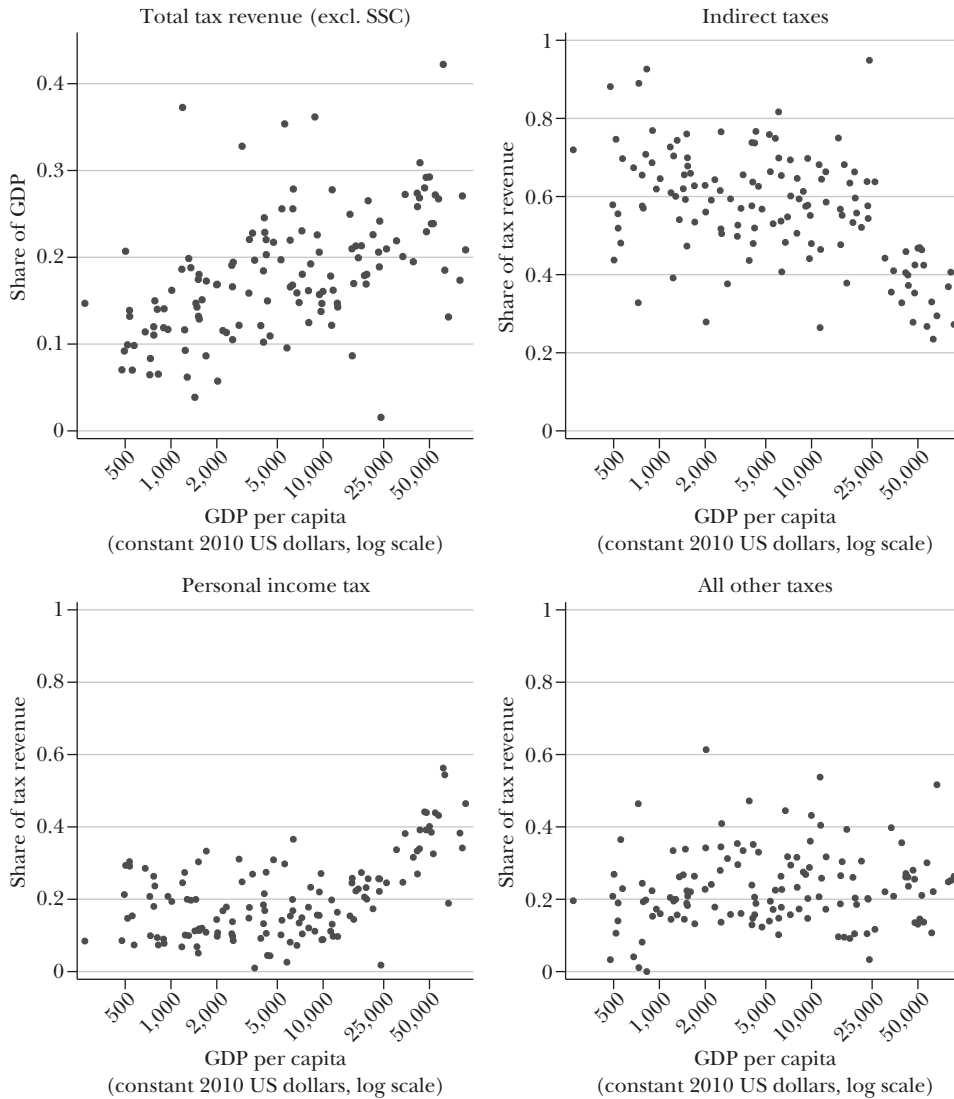
To set the stage for our discussion, Figure 1 plots countries' tax characteristics as a function of their development level, for which we use GDP per capita as a proxy. Table 1 presents averages by country income groups. The data includes the 132 countries with a population above one million inhabitants and whose revenue from oil and gas does not exceed one-third of their GDP (because public finance in countries with a high ratio of oil sales/GDP presents different conceptual issues). These countries account for 93 percent of the world's population.

Panel A considers how the total share of GDP collected in taxes evolves from lower-income to higher-income countries. The measure of total taxes to GDP shown here excludes social security contributions and payroll taxes, which have their own logic outside of redistribution concerns. Total taxes collected, expressed as a share of GDP, increase with economic development. Explaining this evolution, first noted by German economist Adolph Wagner (1835–1917), is beyond the scope of this paper. From an equity perspective, however, we expect governments in low-income countries that collect roughly 10 percent of GDP in taxes to have much less scope to redistribute through taxation than governments in the highest-income countries, where the share of GDP collected in taxes is often closer to 30 percent.

In Panels B, C, and D, these tax ratios are separated into three groups: (1) personal income tax; (2) indirect taxes which include the value-added-tax, sales taxes, excises, and tariffs revenue; (3) all other taxes, whose main components are corporate income taxes, property, and wealth taxes. Two key patterns emerge here.

<sup>3</sup>Shah and Whalley (1991) review earlier work on tax incidence considerations in developing countries.

*Figure 1*  
**Tax Systems across the Development Path**



*Source:* Data for 2018 from Bachas et al. (2022), who collect these from various sources including individual countries' government archives, OECD, and UNU WIDER/ICDT.

*Notes:* Total tax revenues exclude social security contributions (SSC), but include personal income taxes, value-added taxes, other indirect taxes, and all other taxes are as a share of total tax revenues. Sample includes the 132 countries with a population above one million inhabitants and whose revenue from oil and gas does not exceed one-third of GDP (Data on oil and gas production from Ross and Madhavi 2015). The countries represented correspond to 93 percent of the world's population. The tax ratios then are separated into three groups: (1) personal income tax; (2) indirect taxes which include the value-added-tax, sales taxes, excises, and tariffs revenue; (3) all other taxes whose main components are corporate income taxes, property, and wealth taxes.

Table 1

**Tax Systems and Informality by Country Income Group**

	<i>Tax revenue (excl. social security contributions)</i>	<i>Personal income tax share</i>	<i>Indirect tax share</i>	<i>Other taxes share</i>	<i>Self- employment</i>	<i>Informal consumption</i>
Low income	11.0	17.0	64.9	18	81.2	85.7
Lower-middle income	16.7	15.9	59.8	24.3	58.3	64.8
Upper-middle income	19.1	15.3	58.4	26.2	33.8	30.4
High income	22.1	29.8	48.4	21.8	16.1	11.5

*Source:* Each number corresponds to an average across countries in a given income group, using the country income classification defined by the World Bank for 2018.

*Note:* Tax revenues taxes to GDP includes all sources of tax revenue but excludes social security contributions and payroll taxes. The tax ratios then are separated into three groups: (1) personal income tax; (2) indirect taxes which include VAT, sales taxes, excises, and tariffs revenue; (3) all other taxes whose main components are corporate income taxes, property, and wealth taxes. The two proxies for the size of the informal sector are the share of traditional consumption in total consumption and the share of self-employment in the active workforce population. We obtain the tax revenue variables from Bachas et al. (2022), who collect these from various sources including OECD and UNU WIDER/ICDT. These are for the year 2018, for the 132 countries with a population above one million inhabitants and whose revenue from oil and gas does not exceed a third of GDP (data on oil and gas production from Ross and Madhavi 2015). These countries account for 93 percent of the world's population. The data on informal consumption are taken from Bachas, Gadenne, and Jensen (2023a), for 32 LMICs based on expenditure surveys and places of purchase where the consumption occurred. The data on the share of self-employed workers is from the ILO for the same year and set of countries as the tax data. The four income groups are High income countries (HICs), with per capita GDP above \$13,000, Upper-Middle income countries (UMICs), with GDP per capita between \$4,000 and \$13,000, Lower-Middle income countries (LMICs), with GDP per capita between \$1,300 and \$4,000, and Lower-income countries, with GDP per capita below \$1,300.

First, the share of indirect taxes in total taxes decreases with economic development. Low-income countries raise a much larger share of their revenues by means of indirect taxes than high-income countries (65 versus 48 percent).<sup>4</sup> Second, a substitution in the tax mix occurs as per capita GDP rises: the share of indirect taxes declines and the share of personal income taxes rises. This increase in the personal income tax is particularly prevalent at higher levels of economic development. The personal income tax accounts for an average of 16–17 percent of tax revenues in low and lower-middle-income countries; its share of total taxes begins to rise only in upper middle-income countries, reaching 30 percent on average in high-income countries.

Finally, an obvious challenge for tax collection in low- and middle-income countries is that their economies have a substantial informal sector. For our purposes, we shall define informality as the absence of a (full) tax payment by a

<sup>4</sup>Our data does not permit a systematic separation between consumption taxes and taxes on international trade, though it is well-established that trade taxes occupy a smaller share of indirect taxes at increasing levels of development. The decreased reliance on trade taxes is likely to have important equity impacts and this is an interesting area of future research.

firm or individual. It is important to note that, when defined this way, informality arises either because a firm or individual is evading some or all of the taxes they are legally required to pay under the tax code, or because the firm or individual is legally exempt from paying any taxes under the tax code. This point is important in a developing country context because, as we shall see in the following section, vast segments of economic activity are often legally exempt from specific tax bases.

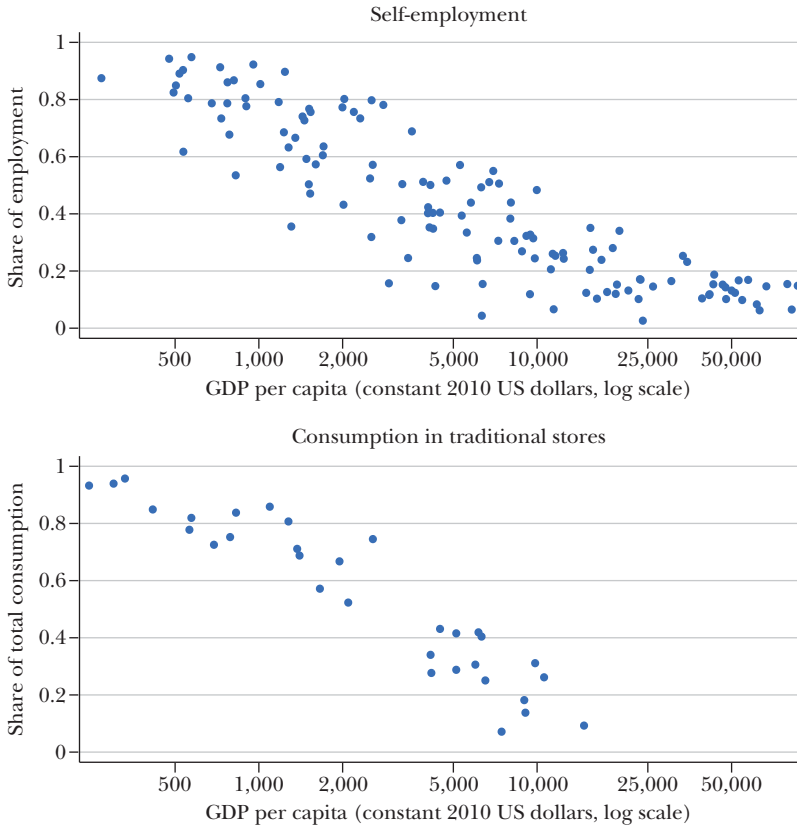
Figure 2 provides two proxy measures of informality at the country level and relates them to economic development. These proxy measures are significant *predictors* of informality. The first measure is the self-employed share in the active workforce. Around the world, including in high-income countries, enforcing income taxes has been found to be more challenging for the self-employed than for employees, due to the absence of third-party reporting and withholding (which, in the case of employees, is done by the employer). The second measure is the share of household consumption from traditional retailers: street stalls and public markets (often called “non-brick and mortar” stores), corner stores, and home production. In comparison to modern retailers (like supermarkets and department stores), traditional retailers are much smaller (in terms of sales and physical space), they hire fewer workers, and have fewer customers and interact with a smaller number of suppliers. All of these characteristics, in turn, are strongly associated with informality, making the share of households’ budget spent in traditional stores a meaningful proxy for informal consumption.<sup>5</sup>

The broad lesson from Figure 2 is that the share of the informal sector in the economy declines steeply as GDP per capita rises. The share of informal consumption decreases from 86 percent in low-income countries to 12 percent in high-income countries; similarly, the share of informal labor decreases from 81 percent in low-income countries to 16 percent in high-income countries. In low- and middle-income countries, a significant challenge for tax design is that a large share of economic transactions and actors operate outside the tax net; in high-income countries, the tax net is much wider.

The patterns in Figure 1 reflect variation in statutory tax incidence across countries, but also differences across settings in *de facto* incidence (through evasion, related to Figure 2) and economic incidence (through market forces which determine the equilibrium values of the tax bases). In the next section, we explain how these patterns and challenges interact to provide perhaps surprising insights into the equity characteristics of taxes in low- and middle-income countries.

<sup>5</sup>Several papers have documented, and provided explanations for, the strong overlap between self-employment and informality (including Kleven et al. 2011; Kleven, Kreiner, and Saez 2016). For more details on the relationship between retailer types and informality, see Bachas, Gadenne, and Jensen (2023a).

Figure 2  
**Informality across the Development Path**



*Source:* The top panel shows the share of self-employed workers using 2020 data from the International Labor Organization for 132 countries (same sample as the one displayed in Figure 1). The bottom panel displays the share of total consumption that occurs in traditional stores (markets, small convenience stores, street stalls, home production) for 32 low- and middle-income countries based on expenditure surveys which record the places of purchase where consumption occurred. The data on informal consumption are taken from Bachas, Gadenne, and Jensen (2023a) and typically correspond to years between 2010 and 2015.

*Note:* This figure displays two proxies for the size of the informal sector across countries at different level of GDP per capita.

### **A Tale of Two Taxes: Personal Income and Value-Added Taxes over the Course of Development**

The tax mix in developing countries is characterized by relatively low levels of personal income tax collection, high reliance on indirect taxes for revenues, and a large share of economic activity occurring in the informal sector. What does this imply for the equity characteristics of two of the most important tax instruments, the

personal income and the value-added-tax? Here, we emphasize two main insights: (1) the existence of large informal sectors explains why low- and middle-income countries levy relatively little in direct taxes via the personal income tax; and (2) the existence of a large informal sector implies that indirect taxes like the value-added tax, which are usually thought of as regressive, instead become progressive.

The first insight is based on observing how both the personal income tax base and employment structure evolve as countries develop. The level of the personal income tax exemption threshold—the income level below which individuals are exempt from paying the tax—moves to a lower position in the country’s income distribution as per capita GDP rises. The decrease in the exemption threshold tracks growth in the employee-share of employment, which occurs gradually further down the country’s income distribution with development. Moreover, whilst the base of the personal income tax expands significantly with development, the share of employees in the tax base remains roughly constant. These facts hold both across countries today and within countries over the long run, as Jensen (2022a) shows with a micro-database of nationally representative household surveys that covers 100 countries at all levels of economic development and long-run time-series in the United States (1870–2010) and Mexico (1960–2010).

These patterns suggest that the expansion of the tax base and the growth in personal income taxation over the development path are driven by a transition from self-employment to employee-employment. These results are consistent with the idea that third-party reporting makes it much easier to enforce taxes on employees than on the self-employed (Kleven, Kreiner, and Saez 2016), so that the structural increase in the employee share as countries develop (Gollin 2008), and the rise of large firms with numerous employees, enables both the growth in personal income taxation and the concomitant decrease in the informal sector. More generally, enforcement constraints—which arise when the self-employed make up a large share of workforce income—shape statutory policy decisions about who to tax. When governments find it challenging to enforce taxes on people with lower incomes, who are mostly self-employed, they react by exempting these large shares of the workforce from the personal income tax altogether.<sup>6</sup>

A narrow tax base places practical limits on the ability of the personal income tax to be a meaningful source of revenue in low- and middle-income countries. This narrowness, likely driven by enforcement constraints, has important implications for the “optimal” way to achieve redistribution, as prescribed by economic theory. Indeed, the most-widely known result in public finance on this question prescribes that all redistribution should be achieved through the personal income tax (Atkinson and Stiglitz 1976)—which implies that redistribution through other tax instruments, such as consumption taxes, is suboptimal. This theoretical result, however, is based on the assumption that governments can implement broad

<sup>6</sup>See Gordon and Li (2009) for a more general discussion of how enforcement considerations constrain governments’ choices of statutory tax instruments. For a discussion of political economy considerations related to broadening the personal income tax base, see Bergolo, Londoño-Vélez, and Tortarolo (2023).



and highly flexible income tax schedules along the full income distribution—an assumption clearly at odds with the reality of most low- and middle-income countries, given their narrow income tax bases. Once we account for the de facto enforcement constraints which shape statutory policy decisions, using consumption taxes for both revenue collection and equity purposes is potentially sound tax policy (Huang and Rios 2016).

The second main insight relates to the equity effects of indirect (consumption) taxes in low- and middle-income countries. Conventional wisdom, when based on the experience of high-income countries, would argue that consumption taxes have null or negative redistributive properties, as they essentially tax households in proportion to their consumption (Warren 2008). This conventional wisdom, combined with the much higher share of indirect taxes and lower share of personal income taxes in total revenues in low- and middle-income countries, may lead one to conclude that the tax system in these countries achieves little or no redistribution.

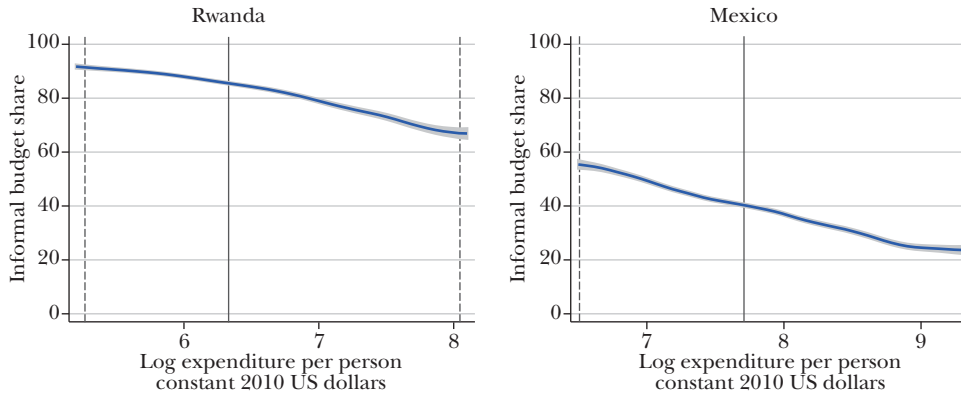
However, taking into account the de facto incidence of consumption taxes can radically change this conclusion.<sup>7</sup> Lower-income households may be more likely to shop in the informal sector; in this case, a consumption tax that effectively only applies to formal consumption may have positive redistributive properties. To investigate whether there is a systematic relationship between the formality status of stores where households shop and their income requires data on the place of purchase for each expenditure. Bachas, Gadenne, and Jensen (2023a) collect and harmonize household expenditure diaries for 32 low- and middle-income countries where they can observe the store-type for each purchase. Motivated by evidence from retail censuses and the literature on informality (for example, Lagakos 2016), they categorize expenditures from modern retailers as formal, and expenditures from traditional retailers, as well as consumption from home production, as informal.

To investigate how shopping patterns relate to household income, Bachas, Gadenne, and Jensen (2023a) use a variation of the Engel curve. A conventional Engel curve shows how an increase in household income is associated with a change in the share of income spent on a specific good. The authors adapt this concept to a setting with widespread informality and establish the existence of what they call the Informality Engel Curve. The Informality Engel Curve shows how a change in household income alters the extent to which a household makes purchases from the informal sector. Figure 3 plots the Informality Engel Curve in two countries: Rwanda and Mexico. In Rwanda, the share of the household budget spent in informal stores falls from 90 percent for the lowest income decile of households to 70 percent for the highest decile. In Mexico, it falls from 55 percent to 25 percent.

The existence of an Informality Engel Curve has several implications for the equity of consumption taxes. First, its downward slope implies that consumption taxes are de facto progressive (under some assumptions on economic incidence to which we will return below). For the average country in Bachas, Gadenne, and

<sup>7</sup>Studies which investigate the distributional impacts of indirect taxes in developing countries include Lustig (2022), Sah (1983), and Jenkins, Jenkins, and Kuo (2006).

Figure 3  
**Informality Engel Curves in Rwanda and Mexico**



*Source:* This figure plots the Informality Engel Curves in Rwanda and in Mexico, copied from Figure 1 in Bachas, Gadenne, and Jensen (2023a).

*Notes:* The line represents the local polynomial fit of the informal budget share (proxied by consumption in traditional stores) on the vertical axis and per person total expenditure on the horizontal axis (measured in log). As households become richer their share of informal consumption tends to decrease. The shaded area around the polynomial fit corresponds to the 95 percent confidence interval. The solid gray line corresponds to the median of each country's expenditure distribution, while the dotted lines correspond to the 5th and 95th percentiles. A description of the data and method used is in the main text.

Jensen (2023a), the magnitude of the slope is sufficiently steep that setting a uniform rate on all formal products is strongly progressive: the effective tax rate paid by the top income decile will be more than twice that of the lowest decile.

Second, consumption of food items and consumption in informal stores has a very strong overlap. Many countries around the world implement reduced consumption tax rates on food products to try to improve equity (motivated by a downward-sloping conventional Engel curve for food). In practice, however, this policy produces little or no additional progressivity in developing countries, because most of the poor households' food consumption is already de facto exempt from taxation since it occurs in traditional, informal stores (and from home production).

Third, while taking into account that Informality Engel Curve increases the equity gains of a consumption tax, it also increases its (distortionary) efficiency costs, relative to a world with no informal consumption. To understand how these countervailing equity and efficiency forces play out and determine the level of the optimal consumption tax rate requires a model set-up (Bachas, Gadenne, and Jensen 2023a). Finally, the optimal extent to which there should be differentiation of consumption tax rates across goods will be limited—both because the equity gains from subsidizing necessity goods relative to other goods is limited, and because such rate-differentiation introduces additional efficiency costs.

Interestingly, in settings where informal consumption and consumption of necessities has an overlap, there may still be a redistributive role for setting reduced tax rates on some necessity goods. Some low- and middle-income countries combine the implementation of ration shops (which sell certain necessity products at a subsidized price) with potential taxation of these same products when they are sold on the market. Combined, these two features (approximately) lead to a piecewise increasing tax schedule. The existence of ration shops can provide households with insurance, because the ration shops permit the option of buying necessity goods at a fixed price in contexts where lack of market integration can lead to large variations in market prices. Implementing this system may be close to an optimal way to redistribute, at least for necessity products typically sold formally, such as energy products (Gadenne 2020). In India, the existence of ration shops has been found to increase welfare significantly, through the insurance gains provided to poorer households (Gadenne et al. 2021).

Underlying our discussion thus far has been an implicit assumption about economic incidence; namely, that there is full pass-through of consumption taxes to prices in formal stores, and there is no pass-through of consumption taxes to prices in informal stores. This assumption supports the baseline result of Bachas, Gadenne, and Jensen (2023a) on the progressivity of consumption taxes in a setting where people with higher income are less likely to purchase from informal stores, but how reasonable is it? To gauge this economic incidence assumption requires data on (tax-inclusive) prices in both formal and informal stores as well as a reform which varies the consumption tax rate. These data requirements are met in Mexico, where the same authors study a reform that increased the rate of the value-added tax in some geographical areas of the country in 2014. Consistent with the baseline assumption, the reform led to a large pass-through of the consumption tax rate increase to prices in formal stores and a much smaller (though nonzero) pass-through to prices in traditional stores.

These results raise several interesting yet unanswered questions. In theory, the coexistence of informal and formal sectors implies that some agents could benefit from a tax increase; for example, a small informal retailer may be able to increase its margins and/or its market share when the formal supermarket next door has to raise its prices in line with a tax rate increase. Brusco and Velayudhan (2023) find evidence consistent with such behavior in India. This possibility holds true not just in retail but in all markets where formal and informal agents interact, including labor markets. Complex patterns of economic incidence will likely change our understanding of equity effects of taxation in low- and middle-income countries. Uncovering these patterns is therefore a promising avenue for future research.

In summary, the existence of large informal sectors in low- and middle-income countries simultaneously limits the redistributive capacity of the personal income tax and makes the indirect consumption tax progressive. Combined, these two insights yield subtle implications regarding the equity effects of policies that seek to increase tax capacity by encouraging more firms or households to formalize. These policies may help governments use the personal income tax to its full

redistributive potential—personal income taxes are, after all, the main redistributive tax instrument used by governments in high-income countries (Jounard, Pisu, and Bloch 2013). However, greater formalization will simultaneously decrease the positive equity effects of consumption taxes, currently a main source of revenue in low- and middle-income countries. Asatryan and Gomstyan (2020) provide some evidence in line with this idea: they show that a reform which improves enforcement of the value-added tax in Armenia led to higher retail prices and had stronger negative effects on low-income consumers.

### **A Bleak House? The Role of Tax Administration, with Property Taxation Examples**

Tax administration is an important tool of policy design in developing countries. However, administrative capacity remains constrained in low- and middle-income countries and governments make frequent reforms to try to alleviate these constraints.<sup>8</sup> While such reforms are usually intended to be distributionally neutral, the practical realities of implementation in the field often create a wedge between the statutory neutrality and the *de facto* incidence impacts. This insight often emerges in settings of local (property) taxation, where administrative reforms can inadvertently have significant distributional consequences due to their effects on tax officials, their use of local information, the extent of their discretion, and the incentive structure they face.<sup>9</sup>

In the local taxation context, local information—information about taxpayers' propensity to pay that is known locally, but not easily observable by outsiders—can be a significant determinant of *de facto* incidence. Some administrative reforms, which initially seek to alleviate capacity constraints, unintentionally stimulate the gathering and use of local information by tax officials in the field and end up having important distributional effects.

Local tax capacity can be constrained by incomplete infrastructure. One important dimension is incomplete street and property addressing infrastructure, which leads tax collectors to struggle to find the intended property in the first place and deliver tax bills. In Ghana, Dzansi et al. (2022) experimentally find that collectors equipped with a GPS tablet to improve navigation in the field delivered

<sup>8</sup>Investments in administrative capacity is a large area of academic work, including Bird and Casanegra de Jantscher (1992) and Besley and Persson (2009). It is also an important area of focus in international organizations; for examples, see the initiatives at TADAT, the Tax Administration Diagnostic Assessment Tool, at <https://www.tadat.org/home>, and ISORA, the International Survey on Revenue Administration, at <https://data.rafit.org/>.

<sup>9</sup>In the local taxation context, research teams in multiple countries have worked in the spirit of what Dufló (2017) called the “plumbing mindset,” basing their work on experiments implemented in collaboration with local governments while focusing on reform details that lie within the feasible scope of tax administrations. See also the paper by Okunogbe and Tourek in this issue for further discussion of tax administrations in developing countries.

27 percent more bills and collected 103 percent more taxes. Even though the tablet was designed (only) to alleviate constraints on bill delivery, collectors used the technology to gather soft information: by leveraging the survey knowledge and time savings induced by navigational improvements, they learned about taxpayers' individual propensity to pay through repeated interactions with property owners and the discovery of hard-to-observe information in the field. In turn, based on this "soft information" gathered on households' income, liquidity constraints, and willingness to comply, the collectors focused on those households with higher propensity to pay. Because income level is one of the main household characteristics which determines propensity to pay, by stimulating the gathering of local information this administrative reform unintentionally had large positive equity effects: the GPS technology strongly increased property tax payments in the top income-asset quartile but had no effect on households in the bottom quartile.

Local information can be built over time, by people who are embedded in communities over prolonged periods. This provides a potentially valuable role for prominent community members to assist in the tax collection process—something that can be achieved through administrative reforms that outsource the responsibility of collecting taxes to local community elites. Local governments in some countries engage local chiefs to participate in the collection of taxes. In the Democratic Republic of Congo, Balán et al. (2022) experimentally find that such outsourcing increases tax collection; moreover, the increase in taxes stems from the local chiefs making use of valuable "soft information" on households' propensity to pay when they decide whom to visit for tax collection. With this type of reform, the equity impacts depend on what type of information tax collectors make use of in the absence of outsourcing. Tax collectors who are outsiders to the local community are likely to rely on more easily observable characteristics, such as exterior house quality, when they decide whom to visit for tax payments. In such a setting, when compared to tax officials who are outsiders, outsourcing tax collection to local chiefs will cause the local tax system to be more regressive in terms of property value (though not necessarily in terms of income, as shown in Balán et al. 2022).

In settings with constrained capacity, including limits on "hard information" such as third-party reports, administrative reforms can create a premium on "soft information" on taxpayers' propensity to pay.<sup>10</sup> The distributional impacts of such administrative reforms are nuanced and hard to predict, because they depend on (1) how the locally relevant dimensions of propensity to pay correlate with observable proxies for equity (for example, either household income or wealth); and (2) how the reform alters the targeting strategies of tax officials in the field. These factors create a wedge between statutory and de facto incidence. Household

<sup>10</sup>Some administrative reforms can also decrease reliance on soft information, as in Okunogbe and Pouliquen (2022). The authors experimentally study the impacts of transitioning taxpayers in Tajikistan from in-person submission of tax returns to electronic filing, finding heterogeneous treatment effects which suggest that the nature of the collection process strongly varied across taxpayers in the pre-reform, more discretionary system.

characteristics, such as propensity to pay, are not readily observable in administrative data and are often only weakly correlated with other taxpayer characteristics that are more easily observable to researchers.<sup>11</sup> Thus, in the design of administrative tax reforms, researchers should seek to gain deeper institutional and contextual knowledge of the local information and tax collector strategies that determine (de facto) tax outcomes in the field. Moreover, how well governments can codify the local, soft information learned in the field by officials and other community members remains an open question.

Reliance on soft information is inherently linked to discretion. Indeed, the absence of regularized processes and extensive monitoring of officials gives tax collectors in the field significant discretion, including over *who* to visit, interact, and target for enforcement and *whose* property to reassess. Due to limited administrative capacity, governments in developing countries may find themselves forced to rely on officials' discretion, at least to some extent—but what are the equity effects of varying the discretion of front-line officials? One prominent area where discretion and equity are linked is where the tax authority has to rely on officials' discretionary assessments in the field to determine the value of the property tax base. In Senegal, Knebelmann, Pouliquen, and Sarr (2023) experimentally find that reducing this form of discretion improves both vertical equity (officials with full discretion tend to undervalue high-value properties) and horizontal equity (officials with discretion display higher variance in estimated property values for properties that have similar true market value).

When discretion is prominent, individual tax officials may resort to their personal distributional preferences to guide them in their activities. In this case, officials' preferences become an important determinant of the wedge between initially designed statutory policies and the de facto incidence in the field.<sup>12</sup> Thus, an important area of future work is to improve the measurement of front-line officials' preferences and integrate them into the evaluation design to study the distributional impacts of administrative reforms.

Administrative reforms that fully remove discretion and soft information are unlikely in the near future in low- and middle-income countries, but some feasible policies can influence the transition towards hard information and regularized processes. This question is particularly relevant with the rapid emergence of new sources of digital data. The distributional effects of integrating digital data will depend on who was benefiting in the absence of systematized observability. In Italy,

<sup>11</sup> Liquidity constraints are likely to be an important determinant of tax compliance in many settings around the world. In Mexico City, Brockmeyer et al. (2021) find that liquidity constraints drive households' responses to changes in local property tax rates. In the presence of liquidity constraints, additional dimensions of tax policy—such as loans for taxpayers—may be helpful to tailor the payment process to household needs, even in the absence of precise individual information. How to design these less-studied dimensions of tax policy is an important area for future research.

<sup>12</sup> Improving the equity of a tax system may increase citizens' willingness to comply with taxes. Evidence on the importance of citizens' distributional preferences is limited. However, the experimental survey evidence across eight low- and middle-income countries in Hoy (2022) is consistent with this hypothesis.

a high-income country setting with traditionally high levels of tax evasion, Rubolino (2023) finds that the integration of satellite imagery into the property registry to expand the local tax base had strong positive equity effects because wealthier homeowners avoided registration in the previous, manual system. Access to more hard information can also improve enforcement capacity, which may, in turn, lead to reforms that alter the statutory tax design. For example, transitioning to an enforcement process that is more regularized and based on “hard data” can potentially allow local governments to implement more statutorily progressive taxes.

However, in settings with limited initial capacity, it is not clear that more data per se will unambiguously reduce the value of discretion (Bachas et al. 2022a). Moreover, it is also not guaranteed that governments with limited capacity can fully remove officials’ discretion (Chalendard et al. 2023). Depending on the nature and quality of the new hard data, coverage of taxpayers based on new data sources may also be uneven. The advent of digital data in government is underway in many low- and middle-income countries and seems unlikely to slow down, but the result is that tax administrations may be headed towards hybrid strategies, in which the use of soft information and discretion varies significantly across segments of taxpayers.

Another important dimension of administrative reform is the incentive structure that governs how tax officials are hired, paid, and promoted.<sup>13</sup> A straightforward model in which payment of taxes and bribes results from bargaining between the tax collector and the taxpayer can generate a rich, if ambiguous, set of predictions where the distributional impacts depend on several factors: the extent to which reformed incentives strengthen the bargaining weights for the tax inspector (or more generally, any uneven strengthening across taxpayer types); the heterogeneity across taxpayers in their disutility from paying bribes or taxes; and the correlation between these dimensions of taxpayer heterogeneity and their income or wealth.

Performance-pay incentive schemes are used in many administrations around the world and provide a good illustration of how incentive reforms can impact the equilibrium bargaining outcome: as the bargaining weight of the tax collector increases, it may impact both the amount of taxes collected and the de facto incidence of who pays (more) taxes. In Pakistan, Khan, Khwaja, and Olken (2016) find that the overall impact of a performance-pay incentive scheme was driven by strong heterogeneity, as incentivized officials collected all the extra property tax revenue from a small group of high-value properties whose tax valuation was revised upwards. Consistent with the incentive reform affecting bargaining weights, the politically connected property-owners were in general less likely to have their valuations revised, but this advantage disappeared when tax collectors had stronger performance incentives.

The bargaining is most often modelled as a setting in which the tax collector can collude with the taxpayer to reduce the amount of taxes paid *in exchange for a bribe*. This highlights the importance of measuring both formal tax payments and

<sup>13</sup>For an in-depth review of the “personnel economics” of the state, see Finan, Olken, and Pande (2017).

“informal” bribe payments, as both matter to determine the ultimate de facto impact on households’ available resources. The incentive reform for tax collectors in Pakistan caused bribes to increase among lower property values where there were no revisions to property valuation—highlighting how tax outcomes are only one part of the de facto payment impacts on households. More generally, almost all the studies cited in this section found significant positive impacts of administrative reforms on *both* tax payments and bribes. For example, the Ghana study found a positive impact on bribes, which was concentrated amongst households in the bottom income-asset quartile (Dzansi et al. 2022).

In summary, even though many administrative reforms are designed with the intention of being implemented uniformly across taxpayers, they end up having significant de facto distributional impacts in practice because of implementation realities in the field. Future work could seek to measure bargaining weights directly, because they contribute significantly to the ultimate impacts of any incentive reform. For example, under a different set of bargaining weights, the same performance-pay scheme used in Pakistan might have led collectors to focus their tax collection efforts on a large number of poorer households that have less influence and spare the small number of more affluent and influential property owners. Given that (partial) reliance on soft information and discretion will likely persist for some time in developing countries, another important question is how to design administrative reforms which retain positive distributional tax impacts while minimizing potentially regressive bribe impacts (Hindricks, Keen, and Muthoo 1999). More work is also required to study rigorously the impact of administrative reforms on *total payments*—the sum of taxes and bribes—at the household level.

Although the discussion in this section has largely focused on local taxation, the insights are likely to carry over to other tax bases that share similar implementation features. Moreover, it is worth noting that formal tax institutions at the local level often coexist with informal institutions that also levy payments on households and firms: informal tax systems, where households make in-kind or in-cash contributions to local public goods (Olken and Singhal 2011; Walker 2022), and social redistributive taxes, which are informal transfers within social and kinship networks (Squires 2021; Carranza et al. 2022). How quantitatively important are these informal systems, what are their distributional impacts, and which equity considerations determine who contributes and how much? These are important questions for future research.

## **Great Expectations: Taxing the Rich in Low- and Middle-Income Countries**

Finding the best way to tax the rich in a globalized world is a major current policy focus globally (Scheuer and Slemrod 2021; Bergolo, Londoño-Vélez, and Tortarolo 2023) and such a step would be highly progressive. Two key issues plague the enforcement of taxes on high-income individuals. First, wealthy individuals earn a substantial share of their income from businesses under their direct or indirect



control (for US-based evidence in this journal, see Kopczuk and Zwick 2020). Taxing business owners is made complex by their ability to allocate income strategically between different categories (such as salaries and profits), to defer taxation, and to utilize business income for direct consumption. To enforce taxes on high income earners, tax administrations require significant audit capacity, ability to link businesses to individuals, and third-party information on all types of earnings—all of which are often limited in low- and middle-income countries.

Second, high-income individuals often locate wealth abroad, particularly in tax haven countries with low tax rates and limited transparency. Financial wealth held offshore in tax havens accounts for an estimated 12 percent of world GDP as of 2022 (Alstadsæter et al. 2023). While not all offshore financial wealth goes unreported, historically the majority does—even in high tax capacity countries. Until recently, international cooperation between tax administrations has been limited (Alstadsæter, Johannesen, and Zucman 2019). Low- and middle-income countries are at least as affected as high-income countries: up to 18 percent of GDP of Africa and the Middle East is held as financial wealth in tax havens, 13 percent in Latin America, and 5 percent in Asia. This compares to 12 percent in Europe and 7 percent in North America (Alstadsæter, Johannesen, and Zucman 2018). As another symptom of the issue of transferring wealth abroad, aid disbursement by the World Bank to the poorest countries in the world is followed by an increase in transactions to tax haven countries, with a leakage rate corresponding to 7 percent of the amount of aid (Andersen, Johannesen, and Rijkers 2021).

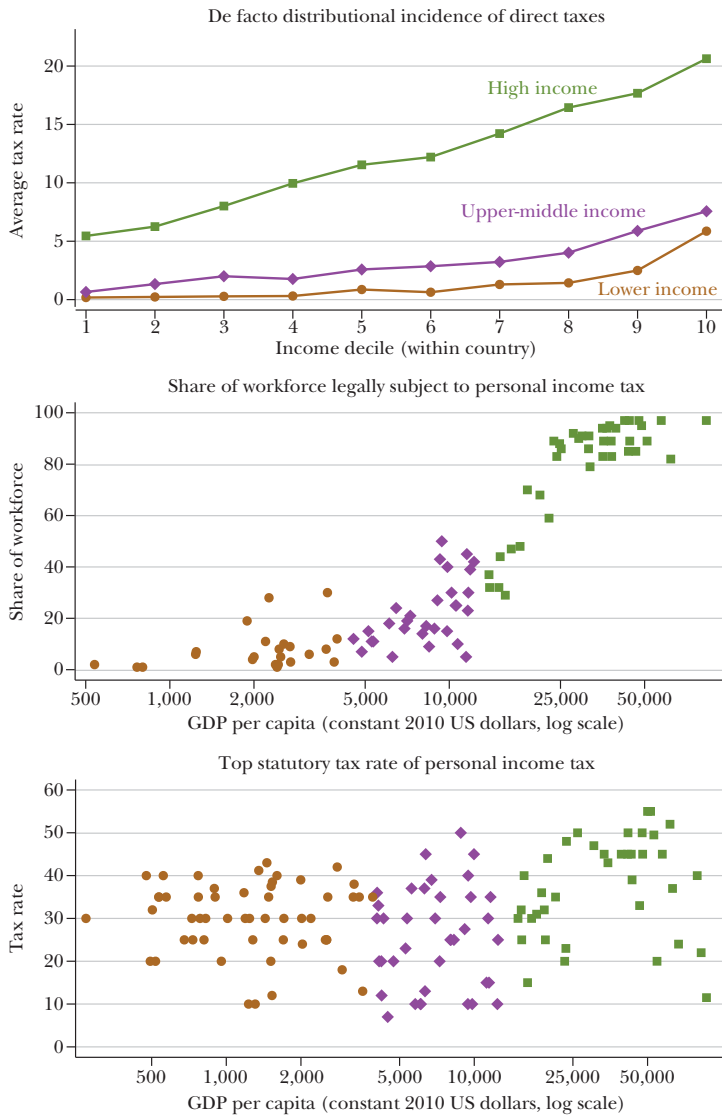
These two challenges—attributing business income to individuals and tracking offshore wealth—severely constrain the progressivity of direct taxes by lowering the de facto tax rates paid by higher-income households. This is true everywhere but is likely to be especially damaging for the progressivity of direct taxation in developing countries.

How, then, does the progressivity of personal income taxation look in practice? While studies of the taxes paid de facto by the very richest (like the top 1 or 0.1 percent) do not exist in a systematic manner, household survey data, combined with some assumptions on taxes paid by salaried workers versus self-employed workers, allows a comparison across countries of the effective income taxes paid by decile of the country's income distribution. For a description of the methodology, see *Commitment to Equity* (Lustig 2022), with data updated for 74 countries by World Bank (2022).

The first panel of Figure 4 plots the share of household budgets that is actually paid in personal income taxes by deciles of the country's income distribution. The categories for countries' level of income follow the World Bank's definition: high-income countries have per capita GDP above \$13,000; upper-middle-income countries have per capita GDP from \$4,000 to \$13,000; and the categories of lower-middle-income and lower-income countries are combined into a single group, with per capita GDP below \$4,000.

The first panel of Figure 4 shows two key patterns that differentiate high-income versus low- and middle-income countries. First, in all deciles, average

Figure 4

**Personal Income Taxation: Statutory and De Facto Incidence**

*Source:* The top panel's data are based on the World Bank (2022) and includes 74 countries for which a fiscal incidence report has been conducted: 30 are high-income countries, 21 are upper-middle-income countries, and 23 are low- and middle-income countries or just low-income. The middle panel shows the share of the active workforce that is covered by the personal income tax, based on the location of the exemption threshold in the income distribution, drawing on data from Jensen (2022a) for 92 countries. The bottom panel shows the top marginal tax rate of the personal income tax using data from 129 countries using 2023 data collected by the authors (and using same sample selection as in Figure 1, more than one million inhabitants and under one-third of GDP from oil and gas).

*Note:* The top panel shows the average income tax rate (de facto) distributional incidence of direct taxation (principally personal income tax) by income deciles separately for High income, Upper-Middle Income and Lower-Middle plus Lower Income countries, as these categories are defined by the World Bank. The fiscal incidence analysis follows the methodology developed by the Commitment to Equity institute (CEQ; see Lustig 2022), which aims to reproduce the de facto incidence of direct taxes using household survey data.

effective tax rates for income taxes are higher in high-income countries. Second, income tax rates in the top deciles increase steeply with development: while the richest 10 percent of households pay around 20 percent of their income in direct taxes in high-income countries, the richest 10 percent of households only pay 5 and 8 percent in low- and in middle-income countries, respectively. The first fact implies that the income tax raises much more revenue in high-income countries, which can later be redistributed via targeted transfers and social insurance programs. The second fact implies that the income tax is much more progressive in high-income countries: the difference in de facto tax rates between the top and bottom decile is around 15 percentage points in these countries, while it is only 5–7 percentage points in low- and middle-income countries.

These patterns occur in part because of the shift from informal to formal employment during the development process, as discussed earlier. To what extent are lower de facto income tax rates also due to statutory policies? On the one hand, the middle panel of Figure 4 shows that the share of the workforce population legally liable to pay personal income taxes significantly increases over development, and only individuals in the top income deciles are liable in low and middle-income countries. This can explain the low levels of de facto tax rates observed in most income deciles apart from at the top. On the other hand, the bottom panel of Figure 4 shows that the top statutory tax rate (the maximum marginal income tax rate that applies to the highest incomes) is only slightly lower in poorer countries: 29 percent on average in low- and middle-income countries versus 35 percent in high-income countries. This difference in top statutory tax rates is clearly insufficient to account for the large difference in de facto income tax rates in the top income decile between high-income countries versus low- and middle-income countries (first panel of Figure 4).

Overall, the large differences between statutory tax rates and de facto tax rates in low- and middle-income countries imply that finding effective ways of taxing individuals with high incomes could have large effects on both revenue and equity. Conversely, taxing the rich may also be particularly difficult in developing countries because of the ample evasion and avoidance opportunities available to high earners, which could lead to large behavioral responses in the form of reductions in reported income when tax rates increase. The size of behavioral responses is typically measured with the elasticity of taxable income, which estimates the percent change in reported income for a 1 percent change in the “net-of-tax rate” (one minus the tax rate).

Only in the last few years has the elasticity of taxable income of rich individuals been estimated in some developing countries, thanks to increased collaborations between tax administrations and researchers. In both Uganda and South Africa, increases in the marginal income tax rate for the 1 percent richest led to large reductions in reported income, with elasticities estimated at just below one (Jouste et al. 2021; Axelson et al. 2023). In Uruguay, an increase of the top income tax rate also led to a substantial drop in reported income (elasticity of 0.6), and to taxpayers switching from filing under the personal income tax to filling under the more

advantageous corporate income tax (Bergolo et al. 2022). Studies of capital taxation in low- and middle-income countries similarly find large behavioral responses: in Colombia, the elasticity of reported wealth to the wealth tax rate is estimated at around two, driven by the underreporting of hard-to-verify assets (Londoño-Vélez and Ávila-Mahecha 2022).

In some cases, changes to reported taxable income can be sufficiently large that raising top tax rates leads to lower total revenue collection (for evidence from Pakistan, see Waseem 2018; for evidence from Brazil, see Locks 2023). More generally, reforms to make the statutory income tax more progressive by increasing tax rates on high earners often lead to large behavioral responses—thereby limiting the effectiveness of these reforms. This likely explains why the personal income tax does not raise as much revenue as in high-income countries, and why the *de facto* income tax rate of top decile earners is limited.

However, elasticities of reported income for top earners are not immutable. Rather, they depend on the availability of tax evasion opportunities and on the enforcement capacity of tax administrations, as recent successful experiences in taxing high-income individuals have shown.<sup>14</sup> In Colombia, a program to disclose hidden wealth voluntarily, in exchange for tax breaks on the wealth disclosed, had a significant effect on tax revenues and progressivity (Londoño-Vélez and Ávila-Mahecha 2021). The disclosure of hidden wealth increased steeply along the wealth distribution, from a probability below 1 percent for the wealthiest 5 percent of households, to a disclosure probability of 40 percent for the wealthiest 0.01 percent. Similarly, the experience of Argentina's 2016 tax amnesty is striking: it led to the revelation of assets worth 21 percent of the country's GDP, with over 80 percent of the assets hidden abroad mainly in the United States and tax havens (Londoño-Vélez and Tortarolo 2022). The effective rate of the wealth tax rose persistently from 0.5 percent to 0.75 percent of wealth, with the largest increase for the top 0.1 percent richest households. Similarly, in Ecuador, a tax on dividend distribution to shareholders located in tax havens encouraged the repatriation of income domestically and raised tax progressivity (Brounstein 2023).

Part of the success of recent tax amnesties can be attributed to the implementation, since 2016, of automatic exchange of information across borders between tax administrations (Alstadsæter et al. 2023). This practice was first started by the United States, under the Foreign Account Tax Compliance Act, and has since been implemented by a majority of countries under the Common Reporting Standard. Under these exchange-of-information agreements, banks report to tax administrations the balances of all accounts held by foreign nationals, and in the case of business accounts, their foreign beneficial owners, too. Third-party information on foreign financial wealth holdings generated by international cooperation has the

<sup>14</sup>Increasing the resources available to tax administrations has been shown to be very effective at reducing tax evasion of the largest taxpayers, and more effective at raising revenue than tax rate hikes: for discussions, see Basri et al. (2021) on enforcement of the corporate tax in Indonesia and Keen and Slemrod (2017) on a theoretical framework for optimal tax administration.

potential to increase revenue and progressivity in low- and middle-income countries and close a prominent channel of tax avoidance used by wealthy individuals. These reforms will likely limit the elasticity of reported income at the top, providing further scope for progressivity.

At the same time, several challenges remain. One important challenge is that real estate wealth is not covered by this exchange of information. This could represent an important constraint on the progressivity of tax systems; in the case of Dubai, real estate wealth held by foreign nationals accounts for several percentage points of the GDP of multiple countries in the region (Alstadsæter et al. 2022). It is unlikely that this real estate wealth is reported domestically; even in Norway, where data could be cross-checked at the taxpayer level, less than one-third of Dubai's Norwegian-owned real estate was reported in the domestic wealth tax. Indeed, wealthy taxpayers are taking advantage of the absence of information exchange for real estate: up to one-quarter of financial wealth in tax havens might have been relocated to real estate following the implementation of information exchanges, where it escapes scrutiny (Bomare and Le Guern Herry 2022). Further, successfully implementing exchange-of-information collaborations presents a set of challenges for countries with low capacity: to receive information they must reciprocate, and the information they receive from foreign banks is not always of sufficient quality to allow their tax administrations to match their citizens to foreign bank accounts.

Finally, an important share of the income of the very wealthy in all countries is tied to business income and undistributed profits, which are typically captured only poorly by personal income tax systems. In this context, the corporate income tax plays an important role as a backstop for the personal income tax (and the progressivity of the tax system as a whole) by ensuring that *some* taxes are collected on business incomes (Fuest and Neumeier 2023). The capacity of the corporate tax to act as a backstop might have eroded over time globally; statutory corporate tax rates have declined and special low tax regimes have flourished, resulting in very large firms paying low effective tax rates. Low- and middle-income countries may be particularly exposed to such trends (Tørsløv, Wier, and Zucman 2022; Johannesen, Tørsløv, and Wier 2020). Registries of beneficial owners of corporations link individuals to the firms they own: in France, these data reveal that the only tax effectively paid by the 75 richest individuals in the country is the corporate income tax (Bach et al. 2023). To our knowledge, similar studies of firm-ownership linkages do not exist in low- and middle-income countries. However, beneficial ownership registries, which identify individuals with significant ownership of a firm, are under development in many countries and will likely provide fruitful research opportunities in the future.<sup>15</sup>

<sup>15</sup> See the Open Ownership website at <https://www.openownership.org/en/map/> for a list of countries that have already enacted or are planning to build a beneficial ownership registry.

## Conclusion

The equity characteristics of tax systems in low- and middle-income countries offer a topic in which the “known unknowns” and “unknown unknowns” outweigh the available “knowns”: many questions remain unanswered, and many questions have not yet been asked. For example, the economic incidence of taxes in the presence of large informal sectors is one important unknown. Insights generated by recent research have challenged some of the received wisdom on the redistributive effects of certain tax instruments, which has been based primarily on the context and research insights from high-income countries.

We emphasize two areas in which policy is rapidly changing in developing countries and that are bound to impact the equity characteristics of tax systems: environmental pricing and the advent of new technologies. While few low- and middle-income countries have implemented a carbon tax and/or environmental pricing schemes, many plans are under consideration. These policy plans are a response to the urgent need to curb future carbon emissions, the bulk of which will come from developing countries (Copeland, Shapiro, and Taylor 2021). Understanding the distributional impacts of environmental fiscal policies across households is key to ensure that such policies are equitable and politically sustainable (Känzig 2023).

In addition, tax administrations in low- and middle-income countries have been making ongoing technology investments to enhance their capacity to raise revenue (Okunogbe and Santoro 2023). Work in this area has mostly focused on the effects of technological changes—such as digitization of payments—on tax collection (for example, Brockmeyer and Sáenz Somarriba 2023; Das et al. 2023). However, the equity effects of new technologies are also likely to be important. A greater reliance on digital technologies to collect taxes may create new forms of horizontal inequities between individuals and firms, depending on their compatibility with technology and their “digital literacy” (Jacobs 2017). The increasing availability of data on taxpayers and their interactions with markets and the government raises the possibility of using new tax instruments that adjust to individual characteristics and circumstances in ways hitherto deemed impossible; an example is a personalized value-added tax, which would compensate poor households on a continuous and individual basis for the amount of taxes paid on their consumption (Kotlikoff, Lagarda, and Marin 2023). The possible gains from improved efficiency and targeting would need to be balanced against data privacy risks and the potential lack of transparency that would arise from more complex tax and transfer systems.

■ *We would like to thank the editors as well as Tim Besley, Mick Keen, Oyebola Okunogbe, Ben Olken, Victor Pouliquen, Enrico Rubolino, Joel Slemrod, Mariano Sosa, Dario Tortarolo, and Matthew Wai-Poi for helpful comments. Pierre Bachas thanks the Norwegian Agency for Development Cooperation (Norad), grant number QZA-22/0011 for funding.*

## References

- Alstadsæter, Annette, Sarah Godar, Panayiotis Nicolaidis, and Gabriel Zucman.** 2023. *Global Tax Evasion Report 2024*. Paris: EU Tax Observatory.
- Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman.** 2018. "Who Owns the Wealth in Tax Havens? Macro Evidence and Implications for Global Inequality." *Journal of Public Economics* 162: 89–100.
- Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman.** 2019. "Tax Evasion and Inequality." *American Economic Review* 109 (6): 2073–2103.
- Alstadsæter, Annette, Bluebery Planterose, Gabriel Zucman, and Andreas Økland.** 2022. "Who Owns Offshore Real Estate? Evidence from Dubai." EU Tax Observatory Working Paper 1.
- Andersen, Jørgen Juel, Niels Johannesen, and Bob Rijkers.** 2021. "Elite Capture of Foreign Aid: Evidence from Offshore Bank Accounts." *Journal of Political Economy* 132 (2): 388–425.
- Asatryan, Zareh, and David Gomtsyan.** 2020. "The Incidence of VAT Evasion." CESifo Working Paper Series 8666.
- Atkinson, Anthony B., and Joseph E. Stiglitz.** 1976. "The Design of Tax Structure: Direct versus Indirect Taxation." *Journal of Public Economics* 6 (1–2): 55–75.
- Axelsson, C., A. Hohmann, J. Pirttilä, R. Raabe, and N. Riedel.** 2023. "What Happens When You Tax the Rich? Evidence from South Africa." Unpublished.
- Bach, L., A. Bozio, A. Guillouzouic, C. Manguyres.** 2023. "Do Billionaires Pay Taxes?" Notes IPP 92 PSE-ESSEC-CREST. [https://www.ipp.eu/wp-content/uploads/2023/06/Note\\_IPP\\_Billionaires-version-actualisee.pdf](https://www.ipp.eu/wp-content/uploads/2023/06/Note_IPP_Billionaires-version-actualisee.pdf).
- Bachas, Pierre, Anne Brockmeyer, Alipio Ferreira, and Bassirou Sarr.** 2022a. "How to Target Enforcement at Scale? Evidence from Tax Audits in Senegal." Economic Development and Institutions Working Paper.
- Bachas, Pierre, Matthew H. Fisher-Post, Anders Jensen, and Gabriel Zucman.** 2022b. "Globalization and Factor Income Taxation." NBER Working Paper 29819.
- Bachas, Pierre, Matthew H. Fisher-Post, Anders Jensen, and Gabriel Zucman.** 2022c. "Database for: Globalization and Factor Income Taxation." NBER Working Paper 29819. <https://globaltaxation.world> (accessed January 1, 2023).
- Bachas, Pierre, Lucie Gadenne, and Anders Jensen.** 2023a. "Informality, Consumption Taxes, and Redistribution." *Review of Economic Studies*. <https://doi.org/10.1093/restud/rdad095>.
- Bachas, Pierre, Lucie Gadenne, and Anders Jensen.** 2023b. "Replication Package for: Informality, Consumption Taxes and Redistribution." Zenodo. <https://doi.org/10.5281/ZENODO.8284123> (accessed January 1, 2023).
- Bachas, Pierre, Anders Jensen, and Lucie Gadenne.** 2024. "Replication data for: Tax Equity in Low- and Middle-Income Countries." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E194851V1>.
- Balán, Pablo, Augustín Bergeron, Gabriel Tourek, and Jonathan L. Weigel.** 2022. "Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the Democratic Republic of the Congo." *American Economic Review* 112 (3): 762–97.
- Basri, M. Chatib, Mayara Felix, Rema Hanna, and Benjamin A. Olken.** 2021. "Tax Administration versus Tax Rates: Evidence from Corporate Taxation in Indonesia." *American Economic Review* 111 (12): 3827–71.
- Bergolo, Marcelo, Gabriel Burdin, Mauricio De Rosa, Matias Giacobasso, Martín Leites, and Horacio Rueda.** 2022. "How Do Top Earners Respond to Taxation? Evidence from a Tax Reform in Uruguay" Unpublished.
- Bergolo, Marcelo, Juliana Londoño-Vélez, and Darío Tortarolo.** 2023. "Tax Progressivity and Taxing the Rich in Developing Countries: Lessons from Latin America." *Oxford Review of Economic Policy* 39 (3): 530–49.
- Besley, Timothy, and Torsten Persson.** 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99 (4): 1218–44.
- Bird, Richard M., and Milka Casanegra de Jantscher, eds.** 1992. *Improving Tax Administrations in Developing Countries*. Washington, DC: International Monetary Fund.
- Boas, H. F., N. Johannesen, C. T. Kreiner, L. Larsen, and G. Zucman.** 2023. "Taxing Capital in a Globalized World: The Effects of Automatic Information Exchange." Unpublished.

- Bomare, Jeanne, and Ségal Le Guern Herry.** 2022. “Will We Ever Be Able to Track Offshore Wealth? Evidence from the Offshore Real Estate Market in the UK.” EUTAX Observatory Working Paper No. 4.
- Brockmeyer, Anne, Alejandro Estefan, Karina Ramírez Arras, and Juan Carlos Suárez Serrato.** 2021. “Taxing Property in Developing Countries: Theory and Evidence from Mexico.” NBER Working Paper 28637.
- Brockmeyer, Anne, and Magaly Sáenz Somarriba.** 2022. “Electronic Payment Technology and Tax Compliance: Evidence from Uruguay’s Financial Inclusion Reform.” World Bank Policy Research Working Paper No. 9947.
- Brusco, Giacomo, and Tejaswi Velayudhan.** 2023. “Does the Informal Sector Escape the VAT?” Unpublished.
- Brounstein, Jakob.** 2023. “Can Countries Unilaterally Mitigate Tax Haven Usage? Evidence from Ecuadorian Transaction Tax Data.” Unpublished.
- Carranza, Eliana, Aletheia Donald, Florian Grosset, and Supreet Kaur.** 2022. “The Social Tax: Redistributive Pressure and Labor Supply.” NBER Working Paper 30438.
- Chalendard, Cyril, Ana M. Fernandes, Gael Raballand, and Bob Rijkers.** 2023. “Corruption in Customs.” *Quarterly Journal of Economics* 138 (1): 575–636.
- Chancel, Lucas, Denis Cogneau, Amory Gethin, Alix Myczkowski, Anne-Sophie Robilliard.** 2023. “Income Inequality in Africa, 1990–2019: Measurement, Patterns, Determinants.” *World Development* 163: 106162.
- Chetty, Raj, Adam Looney, and Kory Kroft.** 2009. “Salience and Taxation: Theory and Evidence.” *American Economic Review* 99 (4): 1145–77.
- Copeland, Brian R., Joseph S. Shapiro, and M. Scott Taylor.** 2021. “Globalization and the Environment.” NBER Working Papers 28797.
- Das, Satadru, Lucie Gadenne, Tushar Nandi, and Ross Warwick.** 2023. “Does Going Cashless Make You Tax-Rich? Evidence from India’s Demonetization Experiment.” *Journal of Public Economics* 224: 104907.
- Duflo, Esther.** 2017. “The Economist as Plumber.” *American Economic Review* 107 (5): 1–26.
- Dzansi, James, Anders Jensen, David Lagakos, and Henry Telli.** 2022. “Technology and Tax Capacity: Evidence from Local Governments in Ghana.” NBER Working Paper 29923.
- Fejerskov Boas, H., N. Johannesen, C. T. Kreiner, L. Larsen, and G. Zucman.** 2023. “Taxing Capital in a Globalized World: The Effects of Automatic Information Exchange.” Unpublished.
- Finan, Frederico, Benjamin A. Olken, and Rohini Pande.** 2017. “The Personnel Economics of the State.” In *Handbook of Field Experiments*, Vol. 2, edited by Abhijit Banerjee and Esther Duflo, 467–514. Amsterdam: North-Holland.
- Fuest, Clemens, and Florian Neumeier.** 2023. “Corporate Taxation.” *Annual Review of Economics* 15: 425–450.
- Gadenne, Lucie.** 2020. “Can Rationing Increase Welfare? Theory and an Application to India’s Ration Shop System.” *American Economic Journal: Economic Policy* 12 (4): 144–77.
- Gadenne, Lucie, Samuel Norris, Monica Singhal, and Sandip Sukhtankar.** 2021. “In-Kind Transfers as Insurance.” NBER Working Paper 28507.
- Gollin, Douglas.** 2008. “Nobody’s Business but my Own: Self-Employment and Small Enterprise in Economic Development.” *Journal of Monetary Economics* 55 (2): 219–33.
- Gordon, Roger, and Wei Li.** 2009. “Tax Structures in Developing Countries: Many Puzzles and a Possible Explanation.” *Journal of Public Economics* 93 (7–8): 855–66.
- Hindriks, Jean, Michael Keen, and Abhinay Muthoo.** 1999. “Corruption, Extortion, and Evasion.” *Journal of Public Economics* 74 (3): 395–430.
- Hohmann, A., J. Pirttilä, R. Raabe, N. Riedel, and C. Axelson.** “What Happens When You Tax the Rich? Evidence from South Africa.”
- Hoy, Christopher.** 2022. “How Does the Progressivity of Taxes and Government Transfers Impact People’s Willingness to Pay? Experimental Evidence across Developing Countries.” World Bank Policy Research Working Paper 10167.
- Huang, Jason, and Juan Rios.** 2016. “Optimal Tax Mix with Income Tax Non-compliance.” *Journal of Public Economics* 144: 52–63.
- Jacobs, Bas.** 2017. “Digitalization and Taxation.” In *Digital Revolutions in Public Finance*, edited by Sanjeev Gupta, Michael Keen, Alpa Shah, and Geneviève Verdier, 25–55. Washington, DC: International Monetary Fund.



- Jenkins, Glenn P., Hatice P. Jenkins, and Chun Yan Kuo. 2006. "Is the Value Added Tax Naturally Progressive?" Unpublished.
- Jensen, Anders. 2022a. "Employment Structure and the Rise of the Modern Tax System." *American Economic Review* 112 (1): 213–34.
- Jensen, Anders. 2021b. "Data and Code for: Employment Structure and the Rise of the Modern Tax System. Nashville, TN: American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E143801V1> (accessed December 22, 2021).
- Johannesen, Niels, Thomas Tørsløv, and Ludvig Wier. 2020. "Are Less Developed Countries More Exposed to Multinational Tax Avoidance? Method and Evidence from Micro-Data." *World Bank Economic Review* 34 (4): 790–809.
- Journard, Isabelle, Mauro Pisu, and Debbie Bloch. 2013. "Tackling Income Inequality: The Role of Taxes and Transfers." *OECD Journal: Economics Studies* 2012 (1): 37–70.
- Jouste, Maria, Tina Kaidu Barugahara, Joseph Okello Ayo, Jukka Pirttilä, and Pia Rattenhuber. 2021. "The Effects of Personal Income Tax Reform on Employees' Taxable Income in Uganda." WIDER Working Paper 2021/11.
- Känzig, Diego R. 2023. "The Unequal Economic Consequences of Carbon Pricing." NBER Working Paper 31221.
- Khan, Adnan Q. Asim I. Khwaja, and Benjamin A. Olken. 2016. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors." *Quarterly Journal of Economics* 131 (1): 219–71.
- Keen, Michael, and Joel Slemrod. 2017. "Optimal Tax Administration." *Journal of Public Economics* 152: 133–42.
- Kotlikoff, Laurence J., Guillermo Lagarda, and Gabriel Marin. 2023. "A Personalized VAT with Capital Transfers: A Reform to Protect Low-Income Households in Mexico." Unpublished.
- Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Pedersen, and Emmanuel Saez. 2011. "Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark." *Econometrica* 79 (3): 651–92.
- Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez. 2016. "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries." *Economica* 83 (330): 219–46.
- Knebelmann, Justine, Victor Pouliquen, and Bassirou Sarr. 2023. "Widening the Tax Net When Information Is Scarce: The Role of Bureaucrats' Discretion." Unpublished.
- Kopczuk, Wojciech, and Eric Zwick. 2020. "Business Incomes at the Top." *Journal of Economic Perspectives* 34 (4): 27–51.
- Lagakos, David. 2016. "Explaining Cross-Country Productivity Differences in Retail Trade." *Journal of Political Economy* 124 (2): 579–620.
- Locks, Gedeão. 2023. "Behavioral Responses to Inheritance Taxes: Evidence from Brazil." Working Paper.
- Londoño-Vélez, Juliana, and Javier Ávila-Mahecha. 2021. "Enforcing Wealth Taxes in the Developing World: Quasi-experimental Evidence from Colombia." *American Economic Review: Insights* 3 (2): 131–48.
- Londoño-Vélez, Juliana, and Javier Ávila-Mahecha. 2022. "Behavioral Responses to Wealth Taxation: Evidence from Colombia." Unpublished.
- Londoño-Vélez, Juliana, and Dario Tortarolo. 2022. "Revealing 21 Per cent of GDP in Hidden Assets: Evidence from Argentina's Tax Amnesties." WIDER Working Paper 2022/103.
- Lustig, Nora, ed. 2022. *Commitment to Equity Handbook: Estimating the Impact of Fiscal Policy on Inequality and Poverty*. Washington, DC: Brookings Institution Press.
- Okunogbe, Oyebola, and Victor Pouliquen. 2022. "Technology, Taxation, and Corruption: Evidence from the Introduction of Electronic Tax Filing." *American Economic Journal: Economic Policy* 14 (1): 341–72.
- Okunogbe, Oyebola, and Fabrizio Santoro. 2023. "The Promise and Limitations of Information Technology for Tax Mobilization." *World Bank Research Observer* 38 (2): 295–324.
- Olken, Benjamin A., and Monica Singhal. 2011. "Informal Taxation." *American Economic Journal: Applied Economics* 3 (4): 1–28.
- Ross, Michael, and Paasha Mahdavi. 2015. "Oil and Gas Data, 1932–2014." Harvard Dataverse. <https://doi.org/10.7910/DVN/ZTPW0Y> (accessed on May 1, 2023).
- Rubolino, Enrico. 2023. "Does Weak Enforcement Deter Tax Progressivity?" *Journal of Public Economics*

- 219: 104833.
- Sah, Raaj Kumar.** 1983. "How Much Redistribution Is Possible through Commodity Taxes?" *Journal of Public Economics* 20 (1): 89–101.
- Scheuer, Florian, and Joel Slemrod.** 2021. "Taxing Our Wealth." *Journal of Economic Perspectives* 35 (1): 207–30.
- Slemrod, Joel.** 2007. "Cheating Ourselves: The Economics of Tax Evasion." *Journal of Economic Perspectives* 21 (1): 25–48.
- Shah, Anwar, and John Whalley.** 1991. "Tax Incidence Analysis of Developing Countries: An Alternative View." *World Bank Economic Review* 5 (3): 535–52.
- Squires, Munir.** 2021. "Kinship Taxation as an Impediment to Growth: Experimental Evidence from Kenyan Microenterprises." Unpublished.
- Tørsløv, Thomas, Ludvig Wier, and Gabriel Zucman.** 2023. "The Missing Profits of Nations." *Review of Economic Studies* 90 (3): 1499–534.
- Walker, Michael.** 2022. "Informal Taxation and Cash Transfers: Experimental Evidence from Kenya." International Growth Centre Working Paper C-89334-KEN-1.
- Warren, Neil.** 2008. "A Review of Studies on the Distributional Impact of Consumption Taxes in OECD Countries." OECD Social, Employment and Migration Working Paper 64.
- Waseem, Mazhar.** 2018. "Taxes, Informality and Income Shifting: Evidence from a Recent Pakistani Tax Reform." *Journal of Public Economics* 157: 41–77.
- World Bank.** 2022. *Poverty and Shared Prosperity Correcting Course*. Washington, DC: World Bank.
- Zucman, Gabriel.** 2013. "The Missing Wealth of Nations: Are Europe and the U.S. Net Debtors or Net Creditors?" *Quarterly Journal of Economics* 128 (3): 1321–64.

# How Can Lower-Income Countries Collect More Taxes? The Role of Technology, Tax Agents, and Politics

Oyebola Okunogbe and Gabriel Tourek

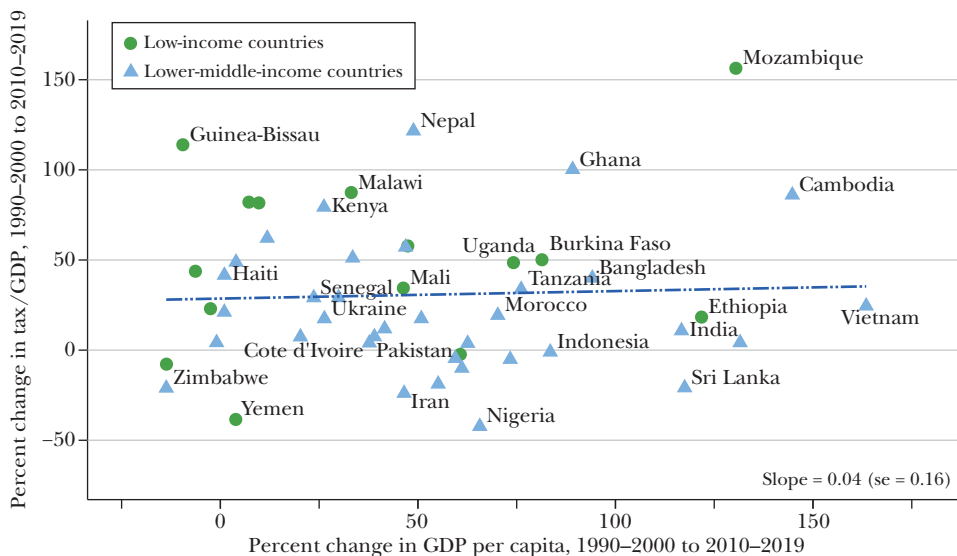
**T**axes are the main source of government revenue in most countries and provide funding for public investments in human capital, infrastructure, and social insurance. Increasing tax revenues is thus a major policy goal in many low- and lower-middle income countries that collect a low share of their GDP as tax revenues. In 2019, for example, the average tax-to-GDP ratio was 12 percent for low-income countries, 18 percent for lower-middle-income countries, 21 percent for upper-middle-income countries, and 30 percent for high-income countries (UNU-WIDER 2022a). Tax-to-GDP ratios among lower-income countries in fact resemble those of modern high-income countries at similar stages of development: among 18 developed countries with available historical data, tax-to-GDP ratios in 1919 averaged 12 percent (Besley and Persson 2014; Mitchell 2007). To account for why developing countries tax so little, Besley and Persson (2014) argued in this journal that low taxation is an outgrowth of deeper economic and institutional factors constraining development, concluding that “the most important challenge is taking steps that encourage development, rather than special measures focused exclusively on improving the tax system.”

This view might suggest that tax capacity should expand with economic development. However, over the last 30 years, the relationship between GDP growth

■ *Oyebola Okunogbe is an Economist in the Development Research Group, World Bank, Washington, DC, and a Research Affiliate at the Centre for Economic Policy Research and the Institute of Fiscal Studies. Gabriel Tourek is Assistant Professor of Economics, University of Pittsburgh, Pittsburgh, Pennsylvania. Their email addresses are [ookunogbe@worldbank.org](mailto:ookunogbe@worldbank.org) and [gabriel.tourek@pitt.edu](mailto:gabriel.tourek@pitt.edu).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.1.81>.

Figure 1  
**GDP per Capita and Tax/GDP, 1990 to 2019**



Source: UNU-WIDER (2022a, b).

Note: This figure displays the relationship between changes in tax revenues as a share of GDP (y-axis) and GDP per capita (x-axis) for low-income and lower-middle-income countries by country income group. Each measure is calculated as the difference between values calculated separately for the periods 1990–2000 and 2010–2019, divided by the value for the base period (1990–2000). We truncate the second period in 2019 to omit the period of the COVID-19 pandemic. We pool annual observations by country within each period as measures for individual years are frequently missing due to intermittent collection of these measures. Values within each period are calculated as the mean of annual observations for each nation within that period. The tax revenues to GDP measure corresponds to tax revenues as a share of GDP within a given year, excluding grants and social security. The income categories are drawn from the World Bank country classifications by income level, as applied in the UNU-WIDER datasets. Low-income countries are displayed as green circles, and lower-middle income countries are displayed as blue triangles. The dashed blue line of best fit is estimated using all displayed observations and has slope 0.04 and robust standard error 0.16.

and taxation levels among low- and lower-middle-income countries has remained essentially flat. Figure 1 plots the relationship between percentage changes from 1990–2000 to 2010–2019 in GDP per capita and tax revenues as a share of GDP for such countries. This pattern suggests economic growth does not automatically generate increases in tax revenue, a point Besley and Persson (2014) also note. Rather, governments must invest in improving the tax system to take advantage of expansions in the tax base that result from increased economic activity.

Improving a tax system requires investments in the ability of the tax administration to carry out its three core tasks: identifying taxpayers, determining how much they owe, and bringing these liabilities into state coffers (Cotton and Dark 2017). Following the framework in Okunogbe (2023), we will refer to

these three dimensions of tax capacity as identification, detection, and collection capacity. Identification capacity is the ability to identify individuals (as well as assets or entities) that are liable to taxation through comprehensive registries and databases. Detection capacity entails verifying the amounts of tax liability, often utilizing data from third parties. Collection capacity is the ability to ensure the payment of liabilities. It involves the use of billing and payment systems to facilitate tax transactions, as well as the imposition of penalties in the face of noncompliance. While policy variables such as tax instruments and tax rates are undoubtedly crucial to the performance of a tax system, this paper focuses on inputs to tax administration. To a large extent, lower-income countries have adopted tax instruments and rates that resemble those in higher-income countries. That low tax revenues persist among the former underscores the value of examining gaps in administrative capacity.

Over the last decade, with the growing availability of high-quality administrative tax data and a rise in partnerships between researchers and governments, there has been a tremendous increase in the amount of evidence on the impact of different interventions to improve tax capacity and increase tax revenues in low- and middle-income countries. This paper examines two major themes from that body of evidence: first, the deployment of new information technology tools to facilitate identification, detection, and collection; and second, the role of tax officials, both in harnessing the potential of technology as well as in complementing it. Despite the availability of viable interventions, an important question is under which conditions governments will choose to invest in tax capacity and expand tax collection. In turn, we will discuss how the current level of taxation, available technology tools, and extent of political competition may affect how the government chooses the nature and level of taxation.

## **The Role of Technology**

Electronic systems that collect, generate, process, and store vast amounts of data can support the implementation of a wide set of tax policy instruments, such as the value-added tax, income tax withholding, customs, and property tax. In lower-income countries, tax modernization programs often seek to support the tax authority in a transition away from relying on manual processes or outdated technology and in-person interactions with taxpayers towards automated, online processes. Of the taxation-related projects approved in 116 countries by the World Bank between 2010 and 2022, 91 countries (78 percent) had a project that included a tax modernization or information technology component, according to our own analysis of World Bank projects. Examples of these projects are integrated tax administration systems for domestic taxes and customs, electronic registration of taxpayers, electronic tax filing and payment, and risk-based audit selection.<sup>1</sup>

<sup>1</sup>Okunogbe and Santoro (2023a) provide a more comprehensive review of the use of technology in tax administration.

A well-integrated information technology system enables the tax authority to identify taxpayers, process information received from taxpayers and third parties to verify liabilities and compliance, and facilitate the remittance of funds to the treasury. We discuss below the use of technology for these three dimensions of tax capacity.

### **Identification Capacity: Taxpayer Identification and Registration**

Identifying taxpayers and developing a taxpayer registry are first-order challenges in many low- and middle-income countries, where proper identification systems often do not exist for individuals nor businesses, properties, or other taxable entities or tax bases. In many countries, people who do not have existing national identification are issued a “Tax ID number” when they enter the tax system. Having a Tax ID that is not linked to an individual’s national identity may serve as a stopgap measure for tax authorities with the most rudimentary systems but may also offer scope for tax avoidance. For example, without a link to personal identity, there is a danger that owners of firms that have accumulated significant tax liabilities could simply shut them down and open new ones to avoid tax penalties. Anecdotally, such maneuvers likely account in part for the high death rates of firms observed in low- and middle-income countries (McKenzie and Paffhausen 2019). The lack of identity systems in many low- and middle-income countries is mirrored in low coverage of personal income taxes, which account on average for 26 percent of tax revenues in high-income countries but only 14 percent in low-income countries (UNU-WIDER 2022a).

Recent technology advances have significantly reduced the costs of identifying individuals at scale and linking identity information across different government functions and registries. For example, countries can enroll millions of people in national systems with biometric identification (like fingerprints or retinal scans) across a nation within a few months.<sup>2</sup> Such identification can then serve purposes not only related to taxation but also to other government functions like verifying eligibility for public benefits or establishing a financial credit history. In the case of real estate taxation, tracking through geographic information systems (GIS) technology can capture property location and features at scale (Knebelmann 2022), facilitating the administration of real estate taxes as well as the provision of public services like postal service, electricity, and sewage.

In an interesting case, Ghana started identifying taxpayers using the “Ghana Card” national identification system in 2021 in place of tax identification numbers issued by the tax authority. Ghanaian tax officials reported that, as a result, they could identify 85 percent of Ghanaians compared to 4 percent under the tax-ID-based system, and the number of filers increased from 4 million to 6.6 million within a few months (GhanaWeb 2022; Ghana News Agency 2021). While this episode does not provide causal evidence, it certainly suggests that an integrated identification

<sup>2</sup>The Kenyan government reportedly registered 36 million citizens for its Huduma Namba digital national ID in the space of two months (Kimani 2019).

system across key sectors of the economy can broaden the reach of the tax authority and may transform a nation's revenue potential.

However, information technologies are not silver bullets. Too often, tax authorities make large investments in taxpayer identification and registration but fail to observe commensurate increases in tax receipts. Means to act on such advances are pivotal. An illustrative case in this vein comes from a randomized experiment in Liberia (Okunogbe 2023). In 2014, the Liberian Revenue Authority undertook a block-by-block digital mapping exercise of a pilot area in the capital, Monrovia, to construct an electronic database of properties and the identities of owners. This database then served as the basis for evaluating distinct tax enforcement measures. Simply informing taxpayers that the tax authority had identified them as the owners of the properties in question (using a notice that included their name and a photo of the property) did not shift the likelihood that tax liabilities were paid (relative to a generic notice). However, combining this identification signal with information about penalties for delinquency more than quadrupled the payment rate. A second experiment sent a signal that the stated penalties would be enforced on defaulters, which achieved a further increase in compliance. This example highlights that tax authorities must complement technology-assisted identification and registration with enforcement actions to translate those investments into tax receipts.

### **Detection Capacity: Third-Party Information and Verification Processes**

Information technology enables a tax authority to verify tax liabilities and detect where evasion might be occurring. In particular, technology facilitates the collection and sharing of third-party information, which is central to modern tax administration. Third-party reporting refers to any information provided to the tax authority about liabilities, other than the declarations taxpayers submit directly. The classic example of third-party information in high-income countries is the information firms provide to tax authorities on employee salaries, an innovation as yet of limited relevance in low- and middle-income countries where over 80 percent of the labor force is employed in the informal economy (Elgin et al. 2021). Even among those employed in the formal sector, firms in low- and middle-income countries often report aggregate employee tax liability without linking amounts to individual employees, who in many cases do not have tax ID numbers (Mayega et al. 2021; Mascagni and Mengistu 2019).

More evidence on how technology allows collection of third-party information in low- and middle-income countries comes from consumption taxes. The value-added tax is increasingly widespread and is one of the most important sources of tax revenue, accounting for on average 28 percent of taxes collected globally (UNU-WIDER 2022a). As Brockmeyer et al. discuss in this issue, a value-added tax has the important built-in feature of collecting third-party information on transactions as firms buy and sell across supply chains. Many countries are moving to digitize this paper trail with the use of electronic fiscal devices, machines that automatically record transactions as they occur and transmit this information to the tax authority through internet or mobile networks. These tools have led to the unveiling of

significant revenue previously undisclosed to the tax authority: in Ethiopia, yielding a 48 percent increase in tax revenue (Mascagni, Mengistu, and Woldeyes 2021). The promise of digital trails does not seem limited to low-income countries, where one might expect more “low hanging fruit” in revenue gains: electronic invoices that made it more difficult to falsify invoices also increased value-added tax revenues in China (Fan et al. 2020), a middle-income country.

Technology can also facilitate digital trails on firms’ revenues when they receive sales payments via credit cards or other electronic payment systems. Even when these transaction records are not automatically transmitted to the tax authorities, the existence of the digital trail may be sufficient to induce a compliance response, as evidenced by the tax impacts resulting from India’s efforts to shift away from paper currency (Das et al. 2023) and the introduction of incentives for credit and debit transactions in Uruguay (Brockmeyer and Sáenz Somarriba 2022). For cross-country evidence, Apeti and Edoh (2023) show that the adoption of mobile money increases total tax collections in developing countries. Third-party information can also come from a wide range of sources, such as utility companies, customs, financial records, and procurement. Sophisticated tax administrations are increasingly linking different databases, facilitated by common and unique identifiers for individuals and firms across sources, to get a more comprehensive picture of true tax liability. Alerting taxpayers to the existence of these data can in turn reduce evasion (for experimental evidence from Costa Rica, see Brockmeyer et al. 2019).

While an increasing number of low- and middle-income countries are taking advantage of information systems to collect third-party information, two important steps are needed to translate information into revenue. First, countries need to use this information systematically in tax enforcement. For example, despite how widespread the value-added tax is, many low- and middle-income countries do not automatically or consistently cross-check invoices and receipts reported by buyers and sellers for the same transactions, which can have sizeable discrepancies. In Uganda, for example, sellers and buyers were found to report different amounts in 79 percent of trading pair-month observations (Almunia et al. 2023). An example of the value of putting third-party information to use comes from Pakistan. After many years of collecting third-party information on firms paying value-added tax but putting the information to limited use, the Pakistani tax authority achieved a 50 percent reduction in value-added tax credits from adopting an automated risk analysis system that leveraged third-party data to reject suspicious credit claims in real time (Shah 2023).

Second, obtaining information on a given tax base from multiple sources is important, too. A consistent pattern arises in which the tax authority attempts to increase enforcement using third-party information obtained on one margin, but taxpayers then adjust on another margin to undo impacts on their overall liability. For example, when tax authorities are given third-party information on revenues, taxpayers may adjust on their expenses or input deductions. Ecuadoran firms notified that they owed more in taxes responded by reporting higher costs (Carrillo, Pomeranz, and Singhal 2017). Ethiopian firms responded the same way when newly



mandated electronic fiscal devices increased firms' reported revenues (Mascagni, Mengistu, and Woldeyes 2021), as did Brazilian firms during an antitax evasion program in Sao Paulo, cutting the initiative's revenue gains in half (Naritomi 2019). Obtaining information on each input to a tax determination will help forestall such adjustments.

Finally, even as countries close potential loopholes within their domestic tax systems, another potential source of revenue loss lies in international tax evasion and avoidance, such as individuals concealing their income or wealth offshore, outside the reach of tax authorities, or multinationals legally shifting their profits to tax havens. These tax losses can be substantial. For example, an amnesty program on wealth taxation in Argentina revealed hidden assets equivalent to 21 percent of GDP (Londoño-Vélez and Tortarolo 2022). Profit shifting can be a major source of revenue loss in lower income countries. Data from the *missingprofits.world* website (based on Wier and Zucman 2022) suggest that 26 percent of corporate taxes in Nigeria are lost due to profit shifting, 13 percent from South Africa, and 6 percent from India.

Advances in computer-based data sharing across countries, such as the Tax Information Exchange Agreement and Automatic Exchange of Information's Common Reporting Standards, provide tools to uncover these attempts at avoidance and evasion as countries can obtain information on the overseas accounts of their citizens. Similarly, in regulating profit shifting, having access to data from other jurisdictions enables tax authorities to compare the price at which a firm wants to transfer an asset among its subsidiaries to a set of comparable transactions to determine whether it complies "arms-length" principles—that is, such transactions should happen under the same conditions (including price) as it would between two separate firms. Low- and middle-income countries face challenges gaining access to these data as signing up to these exchanges requires countries to collect the same information in their own countries, for which they may lack capacity. Even after being granted access, low- and middle-income countries are confronted with the crucial step of using these third-party data effectively, through data analytics that match them to tax records and other in-country data sources to detect evasion and avoidance.

### **Collection Capacity: Electronic Tax Transactions**

Another core application of information technology is the task of processing tax information and payments. For taxpayers, electronic filing and payment options reduce tax compliance costs. For tax administrators, they reduce the risk of data-entry errors and free up personnel who would otherwise be needed for processing paper returns. The generation of tax information and tax withholding at its source by a third-party agent—like employers and financial institutions—is another fiscal innovation that is facilitated by the availability of appropriate data systems. These innovations appear important for tax performance globally, with evidence ranging from individual reforms—like withholding of sales tax by credit card companies in Costa Rica (Brockmeyer and Hernandez 2022)—to cross-country analysis comparing

100 countries at various levels of development (Jensen 2022). Electronic processing of tax information and payments can also provide digitized administrative data to the tax authority that may then support the automation of many tax processes, such as advanced analytics for risk-based audits and targeted enforcement efforts.<sup>3</sup>

When a tax authority more often carries out its duties through e-transactions, and less through in-person interactions, there is also less scope for unsanctioned behaviors like extortion and collusion. At the same time, shifts to digitization may weaken tax officials' ability to monitor or gain contextual knowledge about taxpayers. We return to this tradeoff later in the paper when we discuss interactions between tax officials and technology tools.

One important concern in the use of electronic systems to facilitate compliance is the potential effects on equity. (In this issue, Bachas, Jensen, and Gadenne discuss equity considerations in taxation in lower- and middle-income countries in detail.) On the one hand, electronic systems may help lower-income and less-privileged taxpayers. E-filing, e-payment, and withholding can reduce compliance costs, which are often a larger proportion of income for such taxpayers. Similarly, automated systems that log payments can facilitate installment plans to accommodate liquidity-constrained taxpayers (for evidence from Mexico, see Brockmeyer et al. 2023). On the other hand, evidence across different contexts suggests that taxpayers who are female, rural, less educated, or those heading less-established companies are less likely to use e-services and, where such usage is compulsory, may face higher costs of adoption. There is evidence on this theme from small businesses in Nigeria (Efobi et al. 2019), general use of e-tax services in Rwanda (Santoro et al. 2023), and value-added taxes in Kenya, Tanzania (Eilu 2018), and Rwanda (Mascagni et al. 2023). Electronic systems, coupled with other features of the tax system, can also lead to unintended consequences that disadvantage lower-income taxpayers. For example, if taxes are over-withheld (as is common) and tax filing is optional (perhaps due to exemptions for taxpayers earning below a minimum threshold), then nonfilers will face a higher effective average tax rate, with this impact concentrated among lower-income taxpayers (for evidence from Germany, see Hauck and Wallosek 2023).

### **Cross-Country Evidence on Use of Technology for Identification, Detection, and Collection**

To complement these causally identified studies with evidence from the country level, we zoom out to a cross-country description of the relationship between the extent of technology use and a country's tax performance using data from the OECD Inventory of Tax Technology Initiative's 2023 Global Survey on Digitalisation.

<sup>3</sup>In addition to supporting tax compliance monitoring, data from tax systems also support revenue forecasting and financial reporting. Beyond tax administration applications, information generated for tax purposes can also generate indirect benefits for the design of economic policies more generally. For example, data on firm-to-firm transaction networks can inform about the nature of trading structures, their resilience to financial and climate shocks, and opportunities to diversify supply chains. Likewise, measures of corporate growth—captured through income taxes—can assist in the targeting of industrial policy.

The survey catalogs the tax technologies used in 75 countries, capturing variables that can be mapped to our framework to understand how countries use technology to boost their identification, detection, and collection capacity.<sup>4</sup> For identification-related technologies, the survey collects whether the tax authority requires taxpayers to have digital identification with a unique identity number, and whether this ID is based on government-issued documents and/or biometric information, as well as whether taxpayers can register online. For detection, the survey collects whether the tax authority receives data from third parties, such as trade partners, shares databases with other government bodies, requires taxpayers to submit e-invoices or to maintain online cash registers that directly report their sales to the tax authority, or uses artificial intelligence to conduct risk assessments or detect evasion. For collection capacity, the survey records whether taxpayers can file and/or pay taxes online, and whether they can request an extension or set up a payment arrangement online.

For each dimension of tax capacity, we construct an index that counts how many of the possible technologies a country uses as well as an overall index that sums across the three groups.<sup>5</sup> The horizontal axis of Figure 2 shows the standardized scores for each country (in standard deviation units relative to mean zero). The figure reveals a positive correlation between the extent of technology use (captured by the overall index) and tax revenue as a share of GDP, consistent with the case-specific evidence we discuss above. We examine these relationships using exploratory regressions reported in Appendix Table 1 in the online Appendix and find that this correlation persists while holding constant country income level. Further, the relationship appears driven by technology applications for identification and detection in contrast to the collection index, which does not have a detectable relationship with tax performance. Of course, these relationships are simply correlations and should by no means be interpreted as causal, as many factors can affect both the use of technology as well as tax performance. Nevertheless, the patterns in cross-country data are consistent with findings from well-identified studies at the level of individual countries.

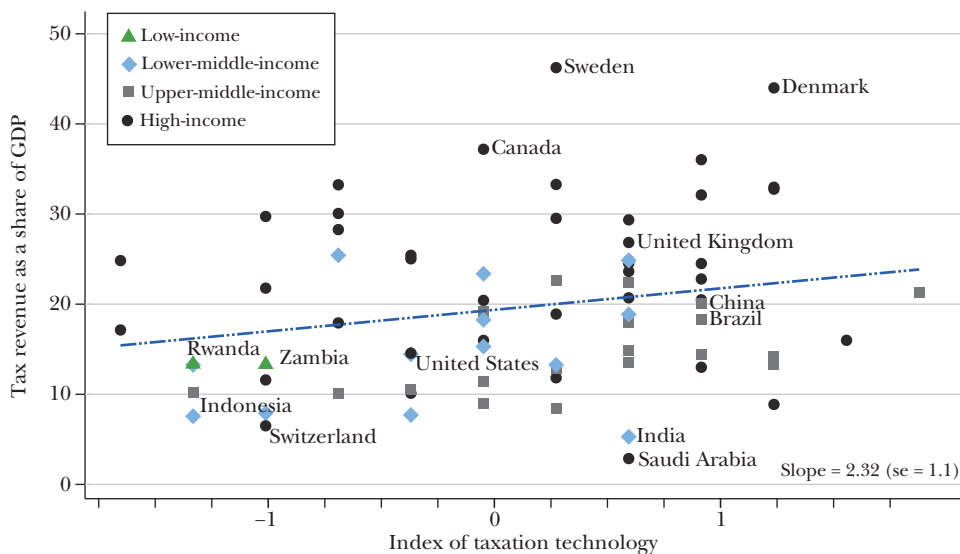
## **The Role of Tax Officials**

Technology can supply data and tools that, ultimately, will assist tax personnel in the work of tax collection. What this work entails depends on the underlying tax system. In more advanced tax systems, tax personnel are deployed mostly in the higher-level tasks of maintaining and updating information technology systems, processing returns, analyzing data, and undertaking audits. In less advanced systems,

<sup>4</sup>For more information, see: <https://www.oecd.org/tax/forum-on-tax-administration/tax-technology-tools-and-digital-solutions/>.

<sup>5</sup>The online Appendix provides further details on the construction of these variables and summary statistics.

Figure 2

**Taxation Technology and Tax-to-GDP**

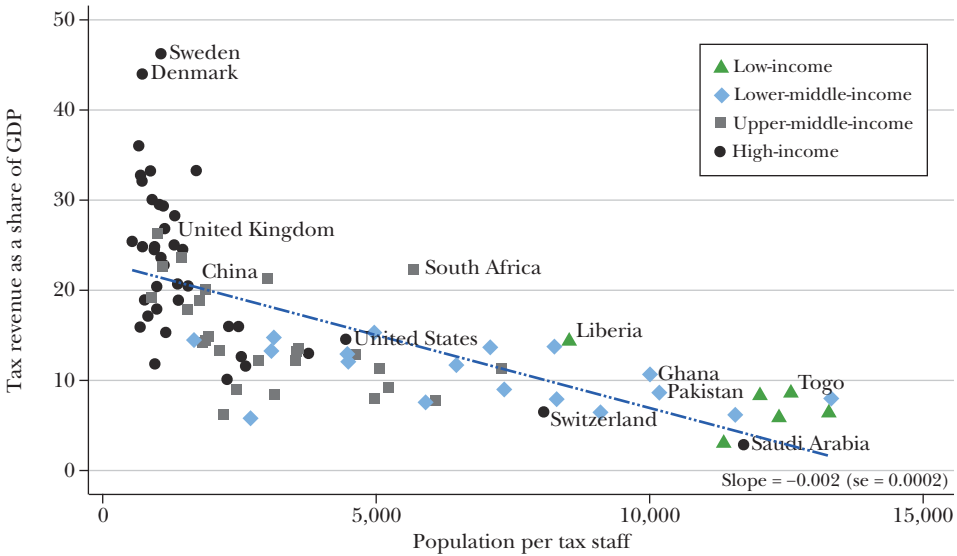
*Source:* Tax revenues as a share of GDP measure drawn from the International Survey on Revenue Administration (CIAT, IMF, IOTA, and OECD 2022), using values for 2018. Taxation technology measure drawn from OECD (2023) Global Survey on Digitalisation.

*Note:* This figure displays the relationship between tax revenues as a share of GDP and an index of taxation technology for 75 individual countries. The tax revenues as a share of GDP measure is drawn from the International Survey on Revenue Administration (CIAT, IMF, IOTA, and OECD 2022) using values for 2018. We fill in missing values for 7 out of 75 countries using UNU-WIDER (2022a) data in 2018. The index of taxation technology is computed from individual measures drawn from the OECD (2023) Global Survey on Digitalisation. The index includes variables capturing tax authorities' use of third-party information, taxpayer identification systems, taxpayer services, and innovation in technology use. More details on these measures are provided in the text and online Appendix. Classifications of country income level are drawn from World Bank Development Indicators. Low-income countries are represented by green triangles, lower-middle-income by blue diamonds, upper-middle by gray squares, and high-income by black circles. The line of best fit is estimated using all observations and has slope 2.32 (robust standard error 1.10).

particularly in the context of local taxation in many lower-income countries, tax collection proceeds manually, with officials soliciting payments and undertaking enforcement efforts in-person.

As a starting point, we examine how the relative strength of a nation's tax staff relates to its fiscal capacity. Figure 3 illustrates a negative relationship between tax-to-GDP and population-to-tax-staff, using data from the International Survey on Revenue Administration (CIAT, IMF, IOTA, and OECD 2022). For ease of illustration, the sample excludes small island nations—which cluster primarily in the low population-to-staff, low tax/GDP quadrant, and where the personnel to tax capacity relationship may fundamentally differ due to reduced scale and their prominence as tax havens—but the observed relationship is robust to including them.

Figure 3  
Population per Tax Staff and Tax-to-GDP



Source: Both measures are drawn from the International Survey on Revenue Administration (CIAT, IMF, IOTA, and OECD 2022), using values for 2018.

Note: This figure displays the relationship between tax revenues as a share of GDP and population per tax staff. Classifications of country income level are drawn from World Bank (2023). Low-income countries are represented by green triangles, lower-middle by blue diamonds, upper-middle by gray squares, and high-income by black circles. The line of best fit is displayed in blue, is estimated using all observations, and has slope -0.002 (robust standard error 0.0002) which translates to an average decrease in the tax-to-GDP ratio of 2 for every additional 1,000 tax personnel.

The ratio of population-to-tax-staff is an order of magnitude smaller in high-income countries. For example, the United Kingdom reported 68,722 full-time-equivalent staff in 2018, or 1,133 citizens per staff, and Sweden 10,486 in total tax staff, or 1,064 citizens per staff. By contrast, middle-income countries like Pakistan (10,176 population per staff) and Ghana (10,006 per staff) and low-income nations like the Democratic Republic of Congo (12,811 population per staff) and Togo (12,582 per staff) possess almost systematically higher ratios. Specialization of tax staff varies with the nature of the economy and the tax system. According to the CIAT, IMF, IOTA, and OECD (2022) database, while the shares of staff assigned to registration, service, and payment activities (at roughly 30 percent) and debt collection (roughly 11 percent) is relatively constant across income groups, high-income countries dedicate a substantially higher share of personnel to audit and investigations (28 percent) than low-income countries (16 percent).

### Deployment of Tax Officials

The effectiveness of tax authorities depends in part on how personnel are deployed in the work of identification, detection, and collection. While strengthening technological capacity may reduce the need for manual functions that can be digitized, the low citizen-to-tax-staff ratios observed in richer countries suggest that personnel remain integral to collection capacity even as economies grow. The specialization of tax staff—and the requisite qualifications—may simply evolve as per capita GDP rises; for example, from in-person registration and monitoring and manual checking of returns to wrangling third-party data to detect evasion.

Another margin for adjustment lies in how personnel are deployed across tax bases, both across the types of taxes a government collects and even the characteristics of taxpayers within bases. In a 2002 policy experiment, Indonesia introduced a corporate tax administration reform that created “medium taxpayer offices” to oversee enforcement of the top several hundred taxpayers in each region. These offices tripled the staff-to-taxpayer ratios assigned to handle such firms. As a result, tax revenue more than doubled, with evidence suggesting this increase derived from more business-reported taxable income and wage bills being reported to the tax authority (Basri et al. 2021). This response implies a sizeable elasticity of collections with respect to the strength and focus of personnel.<sup>6</sup>

As the cross-country pattern in Figure 3 implies, increasing staff—especially in high-return tax functions—could offer a powerful way to ratchet up tax capacity. For example, tax authorities might pursue endeavors with potential to generate large increases in revenue, like concentrating monitoring efforts on large taxpayers or those in evasion-prone sectors. They might also seek to strengthen enforcement in a broad-based way, by training staff to leverage third-party information more effectively. Tax authorities must then also weigh the benefits of such actions against the cost of staff resources that would be expended, as well as costs associated with reallocating personnel away from other functions.

Ultimately, strengthening tax capacity will occur iteratively and must respond to evolving conditions. Yet, improved staff allocation can offer a feasible means of raising collections even absent complementary reforms to tax regimes, administrative processes, or technologies. In an experimental study in Peru, Kapon, Del Carpio, and Chassang (2022) show that “prioritized iterative enforcement” of tax debts—which targets enforcement efforts by trading off expected collection and expected capacity use—can yield collection gains relative to randomly-targeted enforcement efforts. Evidence from the significantly lower capacity setting of the Democratic Republic in the Congo suggests revenue gains as large as 26 percent can be achieved through resource-neutral reallocations of staff (Bergeron, Tourek, and Weigel 2023).<sup>7</sup> The nature of gains

<sup>6</sup>This elasticity can be interpreted as a form the “enforcement elasticity of tax revenue” conceptualized by Keen and Slemrod (2017) as an input to identifying optimal interventions by a tax administration.

<sup>7</sup>Even larger gains may derive from engaging nontraditional personnel in tax collection: in the same setting in the Democratic Republic of Congo, engaging local chiefs in taxation generated a 44 percent

from assignments—in the Congolese case, resulting from high-ability types being more effective when paired with similarly high-ability teammates, particularly when collecting from households with higher payment likelihoods—will likely vary across contexts, but could be fruitfully identified through policy experimentation.

In short, the available evidence is consistent with more staff being able to bring in more revenues, while reallocation of tax officials towards higher return activities can yield nontrivial revenue gains, potentially achieved at low cost, for resource-constrained authorities.

### **Incentives of Tax Officials**

The incentives that tax officials encounter will shape their effectiveness, and by extension, a tax authority's collection capacity. Ample evidence from the public sector suggests incentives can modulate what types of individuals accept roles within bureaucracies and how well they perform. Higher wages or better career benefits offered for public sector positions appear in many cases effective at attracting more able applicants without compromising motivation (in Mexico, Dal Bó, Finan, and Rossi 2013; in Zambia, Ashraf et al. 2020).

In theory, when designing incentive structures, tax authorities should seek to trade off gains in tax collections against the costs of monetary and nonmonetary incentives for staff. In practice, tax administrations operate within government bureaucracies that may constrain options through rigid salary structures, limits on performance incentives, and political considerations. Within the scope of feasible incentive structures, how much can we expect the incentives facing tax officials to matter?

One factor consists in how elastic tax collections are to effort and skills—or other factors like dishonesty—within a certain job. For tax officials engaged in customs, as one example, there might seem less scope for pure effort to impact tax revenues (because of rote processing of total shipments), but larger scope for skills like attentiveness (monitoring product misclassification and underdeclaration to evade tariffs) and dishonesty (bribes in exchange for misclassification) to influence trade tax collections. For service staff tasked with registering informal firms, sheer effort in identifying, locating, and communicating the process of formalization to such businesses may matter comparatively more.

In lower-income countries, the work of taxation still largely involves in-person interactions with taxpayers. On average, 74 percent of firms in low-income countries say they are required to meet with tax officials and on average 3.4 meetings are held per year, compared to 26 percent of firms and 1.8 meetings per year in high-income countries (World Bank Enterprise Surveys). In lower-income settings, tax officials receive payments, often in cash, from individuals and firms, audits require inspection visits to premises rather than examining taxpayer financial accounts,

---

increase in revenues, suggesting that governments with low staff capacity can realize revenue gains by involving agents outside the official state administration (Balán et al. 2022).

and registration drives involving physical outreach are undertaken periodically to extend the tax net over sizeable informal economies. The interpersonal nature of this work—along with officials’ discretion over outcomes—naturally increases opportunities for greater revenue but also corruption.

Consider an example of how these tensions emerge in the lower-middle-income country of Pakistan. In the largest province, Punjab, tax officials are directly responsible for the assessment of property taxes. Tax staff manually construct the property register, including updating for new construction or changes in property values, and wield considerable discretion in determining tax liability through the application of valuation tables and granting of exemptions. At the same time, the wages of these officials are not linked to performance, and few audit mechanisms exist, creating scope for low tax collections and corruption achieved via coercion or collusive agreements between tax officials and taxpayers. This context mirrors the nature of local tax administration in many low- and middle-income countries.

On one hand, tying officials’ incentives more closely to revenue generation could raise collections; on the other hand, raising the return to formal collections could also increase officials’ bargaining power when extracting bribes (in exchange for undervaluation or misclassification) from property-owners. To assess how performance incentives affect tax collections, the Punjab government ran an experiment in collaboration with researchers in 2011 providing bonuses to tax officials in some jurisdictions—in the form of a share of revenue raised for increasing collections (Khan, Khwaja, and Olken 2016).<sup>8</sup> Revenue in jurisdictions with the new performance pay scheme grew by close to 50 percent. The gains derived from a small number of high-value properties being newly taxed at their true value. At the same time, most properties in incentivized areas did not pay more, but reported paying higher bribes. Were properties more homogenously low in value, it is unclear whether the average effect of performance incentives on revenues would have been positive, though the average value of bribes would still likely have increased. A subsequent scheme to incentivize revenue generation through merit-based assignment of postings in Punjab—under which we might expect collusive corruption to be less attractive as a substitute for the reward of obtaining a preferred posting to a new area—resulted in even larger increases in revenue (Khan, Khwaja, and Olken 2019).

The example from Pakistan demonstrates that in settings where scope for collusive corruption is high, the objective function of tax officials should be conceptualized as including their personal gain from wages and formal incentives as well as those from bribes, subject to the costs of being detected. The integration of technology into tax collection may affect this calculus in several ways, as we consider next.

<sup>8</sup> The revenue-based incentive scheme rewarded revenue increases over predetermined benchmarks based on collections in previous years. Bonuses were inversely scaled to the size of jurisdictions to reflect the greater difficulty of increasing revenues in smaller jurisdictions. Additional incentive schemes adjusted performance bonuses based on taxpayer satisfaction and assessment accuracy or evaluations of performance determined by a committee of senior tax officials.



## Interaction of Tax Officials and Technology

Thus far, we have discussed the roles of technology and tax staff as independent forces. In reality, the two may act as substitutes or complements within specific tax functions. On the one hand, technology may almost wholly obviate the need for staff in certain functions (for example, a shift to online tax registration removes the need for manual registration staff), but may also increase the relative return to deploying staff in other roles. For example, the adoption of automatic personal income tax payment through withholding by employers observed in the second half of the twentieth century among high-income nations (Besley and Persson 2014) likely freed up resources previously dedicated to marshaling tax collections through visits, in-person collections, and so on. This shift could account to some extent for the greater shares of staff in audit and investigation roles in higher-income countries relative to lower-income countries we note above.

On the other hand, technology can clearly offer productivity gains for staff by expediting formerly cumbersome processes—like managing tax records or checking returns for errors—and facilitating reallocation of effort towards higher-return activities. The introduction of a digital tax collection technology within a Ghanaian municipality sheds light on this dynamic. In 2021, the municipal tax authority provided a tablet-based technology containing a geospatial database of properties and revenue management software that helped property tax collectors more easily locate properties in the field. A randomly selected subset of local property tax collectors—whose task is to deliver tax bills and collect payments manually—were provided with this new technology. These collectors delivered almost one-third more bills and collected double the amount of revenues compared to collectors persisting in the old, entirely manual system (Dzansi et al. 2022). The timeline of the evolution of visits and payments suggests that the technology assisted collectors in learning more quickly which households were willing and able to make payments and to target their in-person collection efforts accordingly. Technology can thus have high returns in even low resource, low compliance settings.

However, tax authorities face two major challenges. The first is how best to strike a balance between using technology to improve on potential error, bias, or manipulation in the work of human agents, while retaining the value of rich contextual knowledge of an environment that tax officials can acquire from repeated interactions with taxpayers. The tradeoffs involved likely vary with the nature of the task, even within the same setting. As an example, in Senegal, property assessment officials using a “discretionary” method produced real estate valuations with more regressive tax implications than those produced by a computer algorithm (Knebelmann, Pouliquen, and Sarr 2023), while officials charged with identifying firms with high evasion for audit performed better than a newly-introduced algorithm directed toward the same purpose (Bachas et al. 2022).

An example from Tajikistan, in which a set of small and medium-sized businesses were experimentally encouraged to adopt e-filing (Okunogbe and Pouliquen 2022), highlights the tensions around using technology to replace interactions between

taxpayers and officials. Firms that adopted the electronic filing system—replacing a process by which taxes were filed in-person with tax officials—reduced their time spent on tax matters by 40 percent and also freed up the time of tax officials previously tasked with receiving declarations. However, e-filing did not change the average amount of taxes or bribes paid by firms, due to counterbalancing patterns among two types of businesses. One type of firm, identified as being more likely to evade pre-adoption, approximately doubled their amount of taxes paid. It appears that that e-filing disrupted previous collusion between these firms and tax officials that had reduced tax liabilities. For the other type of firms, those deemed less likely to evade, e-filing reduced taxes paid, which is consistent with e-filing leading to less direct monitoring by officials who could have had private information to enforce compliance with firms' true liability.<sup>9</sup> In addition, firms flagged as less likely to evade taxes paid fewer bribes, presumably because e-filing reduced extortion opportunities. This case thus provides a nuanced view of how new technology systems can lead to divergent impacts, depending on the nature of existing relationships and interactions between tax officials and taxpayers.

A second major challenge is that tax officials often resist, manipulate, and in some cases sabotage technology reforms that constrain or monitor their behavior. After all, digitization of information permits greater scrutiny of staff activities, through automatic flagging of errors or data manipulation, generation of reports, and investigation of specific activities (for example, verifying that the outcomes of audits can be replicated with official records). However, in practice, these systems are often difficult to implement successfully, as similar findings among civil servants in other sectors show (for example, Banerjee, Duflo, and Glennerster [2008] documents public sector nurses evading monitoring efforts in India).

A revealing case comes from customs officers in Madagascar. An information technology system was developed to assign customs inspectors randomly to customs declarations, at which import shipments are reviewed and tariffs are assessed. However, in practice, inspectors were able to contravene the official process to assign declarations to particular staff (Chalendar et al. 2023). The manipulation of the assignment process suggests that import brokers bribed staff to be paired with a preferred inspector, who would then facilitate evasion. Upon discovering this scheme, customs management outsourced the assignment process to a third party that used its own software to assign inspectors to declarations randomly. Even then, after a few months inspectors began withholding a portion of the declarations shared with the third party in order to continue to circumvent random assignment. This case drives home the strength of the incentives that tax officials often possess to engage in corruption and the need for bureaucratic reforms to create alternative incentives (for example, by meaningfully raising the cost of being caught).

<sup>9</sup>Some firms reported that this effect resulted from tax officials no longer being able to force them to pay more than their true liability to meet officials' revenue targets.

## **Political Incentives and Constraints on Tax Collection**

Taxation is ultimately the coercion of resources from private citizens to the state, and beyond the tools and personnel of a tax administration, high level political support is arguably the most important determinant of effective taxation. There is substantial evidence from low- and middle-income countries of tax liabilities that are known to the tax authority but go uncollected, despite the availability of relatively low-cost interventions that could be adopted. For example, for property taxes—where the tax base is visible and immobile, and thus presumably easy to tax—collection rates range between 5 and 16 percent across cities in Haiti (Krause 2020), Liberia (Okunogbe 2023), Senegal (Knebelmann 2023), Ghana (Dzansi et al. 2022), Democratic Republic of the Congo (Bergeron, Tourek, and Weigel 2023), and Uganda (Manwaring and Regan 2023). This suggests, in many cases, insufficient political motivation to pursue the collection of these liabilities by enforcing tax laws and imposing financial penalties and legal prosecution on noncompliant taxpayers, some of whom may be economically and politically powerful. A full discussion of the politics of taxation, including bargains with elite taxpayers, reform design to manage resistance from the general population, and politics within the tax administration is beyond the scope of this paper and is extensively discussed in the political science literature (for example, Martin 2023; Moore, Prichard, and Fjeldstad 2018; Prichard 2019; Brautigam, Fjeldstad, and Moore 2008). Instead, we focus the discussion below on evidence relating to factors that determine political choices to invest in tax capacity or to deploy existing capacity.

Broadly, we can consider politicians as weighing the expected benefits and costs of taxation for re-election prospects. On the one hand, higher tax revenues may support re-election by funding the provision of public services, infrastructure investments, social protection, and other public programs that citizens value. On the other, taxing citizens may harm reelection by raising citizen expectations and demands for accountability. Interestingly, this accountability effect can be triggered by simply demanding taxes, regardless of whether those taxes are paid. A study of a door-to-door property tax campaign in Kananga, Democratic Republic of Congo, found higher levels of citizen engagement even among those who were visited by tax authorities but did not comply (Weigel 2020). More broadly, there exists cross-country evidence that tax collection declines prior to competitive elections, suggesting that governments act strategically to avoid negative backlash from citizens (Prichard 2018).

We discuss below three key factors that may shape the cost-benefit calculus of politicians in determining how much taxation to pursue: alternative revenue sources, political competition, and available technologies for reducing the salience and cost of collection. Alternative revenue sources may dampen the impetus to tax. Oil-rich states are a prominent example of a low-tax, low-accountability setting, where oil revenues support a high level of public spending without the need to tax citizens. For instance, oil-rich Persian Gulf countries collect on average 3.6 percent of GDP as taxes (UNU-WIDER 2022a). At the subnational level, transfers from the

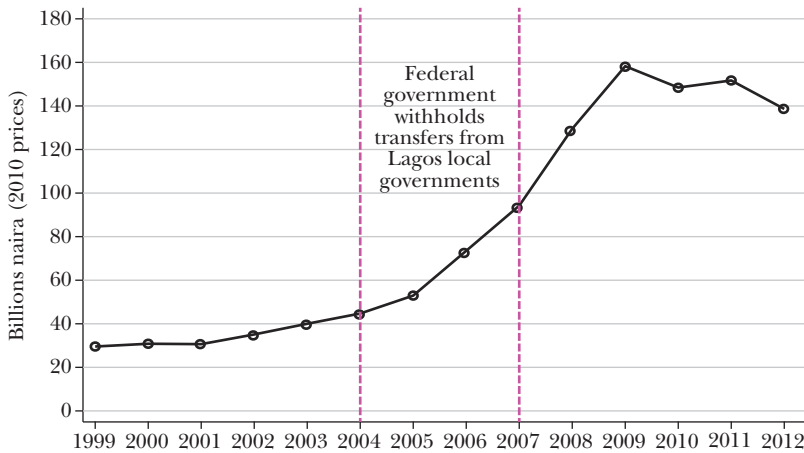
central government may dampen the motivation for local taxation. In Brazil, jurisdictions hit by a negative shock to federal transfers resulting from an update in their population count increased local tax collection by approximately 30 percent (Ferraz, Foremy, and Santini 2023). Localities achieved these expansions of tax collection by broadening the tax base through increased investment in tax capacity, like improving property-tax registers and higher spending on tax agents.

The remarkable expansion of tax collection in Lagos State, Nigeria, offers an interesting historical case indicating the role of political competition, alongside other factors, in generating political support for taxation. Nigeria's government has a strong reliance on oil revenues and correspondingly low levels of nonresource tax collection, with only about 6 percent of GDP collected in nonoil taxes. Following the return to democracy in 1999, Lagos state recorded a dramatic growth in annual tax revenues, shown in Figure 4, from about 30 billion in 1999 to over 150 billion in 2011) in a little over a decade (in Nigerian naira in 2010 prices, this is an increase from about \$200 million to \$1 billion), before the growth tapers off. This revenue growth was greater than in the rest of the country.

Analyses of the Lagos case have highlighted political, economic, and personal factors (de Gramont 2015; Gaspar, Jaramillo, and Wingender 2016; Cheeseman and de Gramont 2017; Bodea and LeBas 2016). The ruling party in Lagos was an opposition party in stiff competition with the central government, and, in the nascent democracy, was keen to show that it could provide a better model of government and public services. Providing a steady source of domestic revenues was thus important not only for meeting the severe needs of road infrastructure, waste management, and security in a congested city, but also in securing funding to consolidate and expand political influence. Although the drive to increase taxation was already underway, the need for tax revenues was underscored in 2004 when the central government withdrew government transfers to Lagos local governments following a dispute over the creation of new local governments. The social conditions in Lagos, with a strong middle class and active civil society, pushed the local government to deliver public goods to win elections, rather than the commonplace strategy of relying on patronage politics. Because Lagos is one of the most populous states and the commercial center of the country, the robust formal sector could support an expansion of personal income taxation through employer withholding. There were also widespread allegations that the leadership of Lagos state held a personal stake in the consulting company that was contracted to process the electronic tax payments, including a lawsuit by the company founder against the former governor of Lagos (Olawoyin 2020). Thus, personal benefit and ability to solidify political power through access to finances may have provided an added incentive.

In Lagos, the specific actions taken by the tax administration focused on deploying technology by introducing electronic payments and tax withholding, by creating an autonomous tax agency separate from the civil service, and by hiring and incentivizing skilled professionals. There was political support for enforcing tax laws, like imposing fines and sealing delinquent businesses. In addition, there was significant taxpayer education and outreach, including campaigns to foster tax

Figure 4  
Tax Revenues in Lagos State, Nigeria



Source: Lagos State Ministry of Economic Planning and Budget (2006, 2010, 2013).

Notes: Adjusted for inflation using the World Bank World Development Indicators Consumer Price Index with 2010 as the base year. Data on tax revenue is from Lagos State Ministry of Economic Planning and Budget (2006, 2010, 2013). The period between the red vertical lines indicates when the federal government withheld transfers from Lagos local governments.

morale by linking the visible public infrastructure investments to the expansion in tax payments. Overall, this case suggests that success in increasing tax collection requires a combination of factors, including a well-motivated government and an ability to deploy diverse tools, particularly those relating to technology and personnel.

Another key factor that may influence a government's decision of how much to enforce taxes is the availability of technology tools that reduce the political and economic costs of taxation by making tax collection more automated and less salient. For example, using employers to withhold personal income taxes removes the need to require citizens to remit payments, and consumption taxes like the value-added tax may be less salient to citizens. Such tools or instruments may lead to less political resistance—but they only work well if the government has access to appropriate technology tools for implementing them. As an example from the US context, Cabral and Hoxby (2012) show that property taxes are less salient to property owners who use the technology-aided practice of bundling their taxes with their monthly mortgage payment through a tax escrow account, compared to other homeowners with mortgages who make one or two large direct tax payments a year to the government. Areas with higher rates of escrow use also have higher property tax rates, suggesting that the lower salience and automation of collection allows those jurisdictions to collect more revenues.

One contemporary example of how the availability of technology can influence the tax choices of governments in low- and middle-income countries is the emergence of taxes on mobile money transactions in a number of African countries. Digital payments via mobile phones have seen substantial growth in the region, serving as a major driver of financial inclusion (Suri and Jack 2016). For governments, a major justification for taxes on mobile money transactions is to expand the tax net over large informal sectors that are traditionally outside the current reach of tax authorities. A mobile money tax is relatively easy to monitor and can be collected by telecommunication companies and service providers, without the need for a tax authority to make costly investments in identification, detection, and collection capacity. Critics have called it a “lazy tax” (Karombo 2022). Tellingly, the salience of the tax has led to stiff public opposition, with some countries having to reduce the tax rate (for example, Ghana, Tanzania, Uganda) and others to completely abandon it (like Malawi).

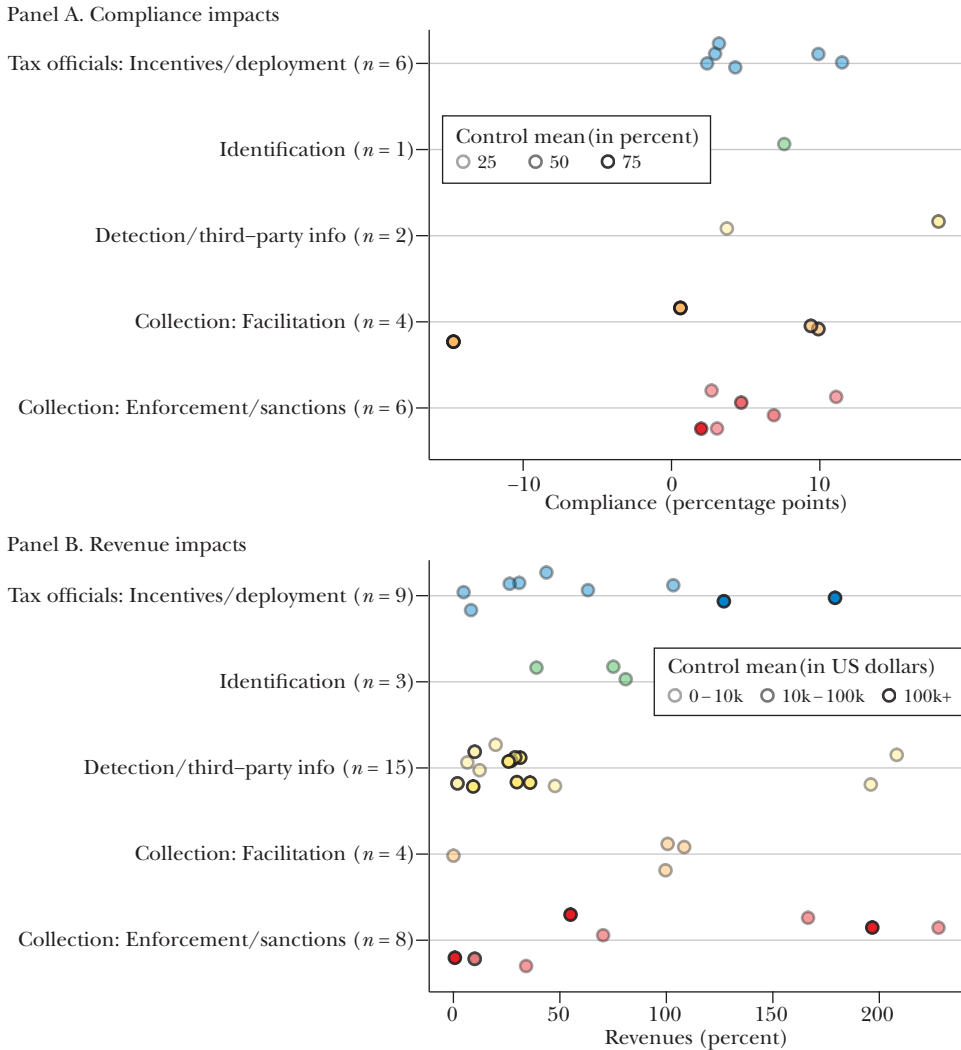
Finally, the decision of whether or not to pursue investments in tax capacity may depend on the current levels of tax collection. In settings with low levels of tax collection, initial investments in tax capacity—such as building a tax database or establishing technology for tax monitoring—typically have high financial costs and may yield only marginal revenue gains initially. As a result, at the early stages of trying to expand taxation, governments in this position can be caught in a bind: they do not yet have sufficient resources to significantly expand public good provision to respond to the increased demands, public scrutiny, and backlash that strengthening tax capacity might trigger among citizens. As such, political factors may contribute to the persistence of low tax-low accountability traps (Besley and Persson 2009).

## Discussion

Transforming a nation’s tax capacity requires intentional investments in the development of tax systems. Based on the growing body of evidence on this topic, we have highlighted the role that two factors—information technology and tax officials—can play in this process in low- and middle-income countries.

We have discussed many of the relevant studies, and in Figure 5, we attempt to summarize recent work showing the range impacts from successful interventions. The figure displays effect sizes observed in experimental or quasi-experimental studies published since 2011 that examine tax interventions in low- and middle-income countries. We include five types of interventions that map to our framework and preceding sections: (1) tax officials-incentives/deployment are interventions in these areas; (2) identification are interventions that help to identify or register taxpayers; (3) detection/third-party information refers to interventions that provide information on the liabilities of taxpayers; (4) collection-facilitation are interventions that make it easier to pay taxes; and (5) collection-enforcement are interventions focused on the penalties for noncompliance. Panel A shows effects on the extensive margin of compliance, the percentage of taxpayers that pay the

Figure 5  
Impacts of Tax Interventions



Source: The full list of studies and results compiled by the authors is available in the online Appendix.  
 Note: This figure displays the effect sizes of tax interventions from existing studies. The full list of studies and corresponding details can be found in the replication materials. The y-axis displays the type of interventions considered, followed by the number of effect sizes included for each type (denoted by n). For Panel A, the x-axis displays the change in tax compliance in percentage points. For Panel B, the x-axis displays changes in revenues in percentages. The opacity of encircled dots is relative to the size of the control means (denoted in percentages for Panel A, and in US dollars for Panel B)—that is, darker dots indicate higher value control means. In both the panels, each intervention type is captured by a different color of the encircled dots. Panel A includes 19 interventions compiled from 13 studies, and Panel B includes 39 interventions compiled from 26 studies.

tax, with effects measured in percentage point changes relative to the mean of the comparison group. It includes 19 interventions across 13 studies. Panel B shows the gains in tax revenues, expressed as a percent of the comparison sample mean. It includes 39 separate interventions across 26 different studies.

As the goal is to show the relative effect sizes of successful interventions, in both panels we include only headline effects that are statistically significant at the 10 percent level or lower. In addition, because low initial levels may magnify gains when expressed in percentage terms, darker dots indicate higher values of comparison means; that is, a given change is judged more substantial if achieved relative to a larger reference level.

The studies summarized in Figure 5, as well as the broader case study evidence discussed throughout the paper, suggest a wide menu of tax administration choices available to low- and middle-income countries. In panel A, effect sizes are broadly similar across intervention categories, suggesting that extensive margin compliance can be increased from a variety of inputs to strengthening capacity.<sup>10</sup> Panel B shows that the largest revenue gains observed within this set of studies come from interventions aimed at improving collection via enforcement, deployment of tax officials (the largest gains observed in this category are from studies of deployment rather than incentives), and improving detection via third-party information. In the most striking cases, revenues have more than doubled or even tripled, even at substantial levels of existing tax collection.

Of course, in interpreting the results in Figure 5, it is important to remember that many of these studies were done in local areas, not entire countries, and cover a range of different taxes. The interventions considered by this body of work are also a function of where research has thus far been feasible and of where the interests of researchers have aligned with those of policymakers. Thus, both the evidence summarized in Figure 5 and the case studies we have called attention to in this paper are not meant to suggest a comprehensive nor “one size fits all” approach to tax administration choices for developing countries. After all, deliberations about tax mobilization will inevitably confront an array of issues, both economic and political. Instead, we intend our discussion to offer inspiration—and caveats—for governments and tax authorities considering the value of technologies and what implications their use carries for the direction of tax personnel.

There remains much to learn beyond our current understanding of how these factors contribute to tax capacity. One clear area for future exploration is understanding how to take advantage of potential complementarities between technology and tax officials, where staff recruitment and the skills will likely play important roles. Moreover, technological advances will undoubtedly pose evolving challenges for taxation, including how to tax digital activities and incorporate digital currencies. Finally, given how fundamental political support is for deploying tax capacity, it

<sup>10</sup>The outlier, with a negative impact, is the Tajikistan e-filing result (Okunogbe and Pouliquen 2022) discussed previously, which likely resulted from reduced scrutiny by officials.



is imperative to build our understanding of the factors that shape how governments choose when, how, and how much to tax.

■ We thank Archita Misra and Kairat Umargaliyev for remarkable research assistance. We are grateful to Pierre Bachas, Anne Brockmeyer, Matthew Collin, Anders Jensen, Wilson Prichard, and Mahvish Shaukat for feedback and helpful conversations.

## References

- Almunia, Miguel, Jonas Hjort, Justine Knebelmann, and Lin Tian. 2023. "Strategic or Confused Firms? Evidence from 'Missing' Transactions in Uganda." *Review of Economics and Statistics*. [https://doi.org/10.1162/rest\\_a\\_01180](https://doi.org/10.1162/rest_a_01180).
- Apeti, Ablam E. and Eyah D. 2023. "Tax revenue and mobile money in developing countries." *Journal of Development Economics* 163: 103014.
- Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S. Lee. 2020. "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services." *American Economic Review* 110 (5): 1355–94.
- Bachas, Pierre, Anne Brockmeyer, Alipio Ferreira, and Bassirou Sarr. 2022. "How to Target Enforcement at Scale? Evidence from Tax Audits in Senegal." Economic Development and Institutions Working Paper.
- Bachas, Pierre, Lucie Gadenne, and Anders Jensen. Forthcoming. "The Equity of Tax Systems in Low- and Middle-Income Countries." *Journal of Economic Perspectives*.
- Balán, Pablo, Augustin Bergeron, Gabriel Tourek, and Jonathan L. Weigel. 2022. "Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the Democratic Republic of the Congo." *American Economic Review* 112 (3): 762–97.
- Banerjee, Abhijit V., Esther Duflo, and Rachel Glennerster. 2008. "Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System." *Journal of the European Economic Association* 6 (2-3): 487–500.
- Basri, M. Chatib, Mayara Felix, Rema Hanna, and Benjamin A. Olken. 2021. "Tax Administration versus Tax Rates: Evidence from Corporate Taxation in Indonesia." *American Economic Review* 111 (12): 3827–71.
- Bergeron, Augustin, Gabriel Z. Tourek, and Jonathan L. Weigel. 2023. "The State Capacity Ceiling on Tax Rates: Evidence from Randomized Tax Abatements in the DRC." NBER Working Paper 31685.
- Besley, Timothy, and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99 (4): 1218–44.
- Besley, Timothy, and Torsten Persson. 2014. "Why Do Developing Countries Tax So Little?" *Journal of Economic Perspectives* 28 (4): 99–120.
- Bodea, Cristina, and Adrienne LeBas. 2016. "The Origins of Voluntary Compliance: Attitudes toward Taxation in Urban Nigeria." *British Journal of Political Science* 46 (1): 215–38.
- Brautigam, Deborah, Odd-Helge Fjeldstad, and Mick Moore, eds. 2008. *Taxation and State-Building in Developing Countries: Capacity and Consent*. Cambridge, UK: Cambridge University Press.
- Brockmeyer, Anne, Alejandro Estefan, Karina Ramírez Arras, and Juan Carlos Suárez Serrato. 2023. "Taxing Property in Developing Countries: Theory and Evidence from Mexico." NBER Working Paper 28637.
- Brockmeyer, Anne, and Marco Hernandez. 2022. "Taxation, Information, and Withholding: Evidence from Costa Rica." CEPR Discussion Paper 17716.

- Brockmeyer, Anne, and Magaly Sáenz Somarriba.** 2022. “Electronic Payment Technology and Tax Compliance: Evidence from Uruguay’s Financial Inclusion Reform.” CEPR Discussion Paper 17097.
- Brockmeyer, Anne, Spencer Smith, Marco Hernandez, and Stewart Kettle.** 2019. “Casting a Wider Tax Net: Experimental Evidence from Costa Rica.” *American Economic Journal: Economic Policy* 11 (3): 55–87.
- Cabral, Marika, and Caroline Hoxby.** 2012. “The Hated Property Tax: Salience, Tax Rates, and Tax Revolts.” NBER Working Paper 18514.
- Carrillo, Paul, Dina Pomeranz, and Monica Singhal.** 2017. “Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement.” *American Economic Journal: Applied Economics* 9 (2): 144–64.
- Chalendard, Cyril, Ana M. Fernandes, Gael Raballand, and Bob Rijkers.** 2023. “Corruption in Customs.” *Quarterly Journal of Economics* 138 (1): 575–636.
- Cheeseman, Nic, and Diane de Gramont.** 2017. “Managing a Mega-city: Learning the Lessons from Lagos.” *Oxford Review of Economic Policy* 33 (3): 457–77.
- CIAT, IMF, IOTA, and OECD.** 2022. “International Survey on Revenue Administration: 2020 and 2021.” <https://data.rafit.org/?sk=8b008788-ebde-4d61-bc90-7438d6aa12dc&Id=1637191076670> (accessed on June 2, 2023)..
- Cotton, Margaret, and Gregory Dark.** 2017. *Use of Technology in Tax Administrations 2: Core Information Technology Systems in Tax Administrations*. Washington, DC: International Monetary Fund.
- Dal Bó, Ernesto, Frederico Finan, and Martín A. Rossi.** 2013. “Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service.” *Quarterly Journal of Economics* 128 (3): 1169–1218.
- Das, Satadru, Lucie Gadenne, Tushar Nandi, and Ross Warwick.** 2023. “Does Going Cashless Make You Tax-Rich? Evidence from India’s Demonetization Experiment.” *Journal of Public Economics* 224: 104907.
- de Gramont, Diane.** 2015. “Governing Lagos: Unlocking the Politics of Reform.” Carnegie Endowment for International Peace, January 12. <https://carnegieendowment.org/2015/01/12/governing-lagos-unlocking-politics-of-reform-pub-57671>.
- Dzansi, James, Anders Jensen, David Lagakos, and Henry Telli.** 2022. “Technology and Tax Capacity: Evidence from Ghana.” NBER Working Paper 29923.
- Efobi, Uchenna, Ibukun Beecroft, Tanankem Belmondo, and Amelia Katan.** 2019. “Small Business Use of the Integrated Tax Administration System in Nigeria.” International Centre for Tax and Development African Tax Administration Paper 8.
- Eilu, Emmanuel.** 2018. “Adoption of Electronic Fiscal Devices (EFDs) for Value-Added Tax (VAT) Collection in Kenya and Tanzania: A Systematic Review.” *African Journal of Information and Communication* 22: 111–34.
- Elgin, Ceyhan, M. Ayhan Kose, Franziska Ohnsorge, and Shu Yu.** 2021. “Understanding Informality.” CEPR Discussion Paper 16497.
- Fan, Haichao, Yu Liu, Nancy Qian, and Jaya Wen.** 2020. “Computerizing VAT Invoices in China.” NBER Working Paper 24414.
- Ferraz, Claudio, Dirk Foremy, and Juan Francisco Santini.** 2023. “Revenue Slumps and Fiscal Capacity: Evidence from Brazil.” Unpublished.
- Gaspar, Vitor, Laura Jaramillo, and Philippe Wingender.** 2016. “Political Institutions, State Building, and Tax Capacity: Crossing the Tipping Point.” International Monetary Fund Working Paper 16/233.
- Ghana News Agency.** 2021. “GRA Identifies over 14 million Tax Payers after Integration of Ghana Card/TIN.” July 22. <https://gna.org.gh/2021/07/gra-identifies-over-14-million-tax-payers-after-integration-of-ghana-cardtin> (accessed July 14, 2023).
- GhanaWeb.** 2022. “Ghana Card Shoots Up Ghanaians with TIN from 4% to 85%—Bawumia.” July 20. <https://www.ghanaweb.com/GhanaHomePage/business/Ghana-card-shoots-up-Ghanaians-with-TIN-from-4-to-85-Bawumia-1586516> (accessed July 14, 2023).
- Hauck, Tobias, and Luisa Wallossek.** 2023. “Optional (Non-)filing and Effective Taxation.” Unpublished.
- Jensen, Anders.** 2022. “Employment Structure and the Rise of the Modern Tax System.” *American Economic Review* 112 (1): 213–34.
- Kapon, Samuel, Lucia Del Carpio, and Sylvain Chassang.** 2022. “Using Divide-and-Conquer to Improve Tax Collection.” NBER Working Paper 30218.
- Karombo, Tawanda.** 2022. “It’s a Lazy Tax’: Why African Governments’ Obsession with Mobile Money Could Backfire.” *Rest of World*, January 7. <https://restofworld.org/2022/how-mobile-money-became-the-new-cash-cow-for-african-governments-but-at-a-cost/>.

- Keen, Michael, and Joel Slemrod.** 2017. "Optimal Tax Administration." *Journal of Public Economics* 152: 133–42.
- Khan, Adnan Q., Asim I. Khwaja, and Benjamin A. Olken.** 2016. "Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors." *Quarterly Journal of Economics* 131 (1): 219–71.
- Khan, Adnan Q., Asim Ijaz Khwaja, and Benjamin A. Olken.** 2019. "Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings." *American Economic Review* 109 (1): 237–70.
- Kimani, Trizza.** 2019. "36 million Registered for Huduma Namba." *The Star*, May 24. <https://www.the-star.co.ke/news/2019-05-24-36-million-registered-for-huduma-namba> (accessed November 2023).
- Knebelmann, Justine.** 2022. *Digitalisation of Property Taxation in Developing Countries: Recent Advances and Remaining Challenges*. London: Overseas Development Institute.
- Knebelmann, Justine, Victor Pouliquen, and Bassirou Sarr.** 2023. "Discretion versus Algorithms: Bureaucrats and Tax Equity in Senegal." Unpublished.
- Krause, Benjamin.** 2020. "Balancing Purse and Peace: Tax Collection, Public Goods and Protests." Working Paper.
- Lagos State Ministry of Economic Planning and Budget.** 2006. "Digest of Statistics 2006." <https://mepb.lagosstate.gov.ng/lbs-publication> (accessed November 7, 2023).
- Lagos State Ministry of Economic Planning and Budget.** 2010. "Digest of Statistics 2010." <https://mepb.lagosstate.gov.ng/lbs-publication> (accessed November 7, 2023).
- Lagos State Ministry of Economic Planning and Budget.** 2013. "Digest of Statistics 2013." <https://mepb.lagosstate.gov.ng/lbs-publication> (accessed November 7, 2023).
- Londoño-Vélez, Juliana, and Dario Tortarolo.** 2022. "Revealing 21 per cent of GDP in Hidden Assets: Evidence from Argentina's Tax Amnesties." UNU-WIDER Working Paper 2022/103.
- Martin, Lucy E. S.** 2023. *Strategic Taxation: Fiscal Capacity and Accountability in African States*. New York: Oxford University Press.
- Mascagni, Giulia, Roel Dom, Fabrizio Santoro, and Denis Mukama.** 2023. "The VAT in Practice: Equity, Enforcement, and Complexity." *International Tax and Public Finance* 30 (2): 525–63.
- Mascagni, Giulia, and Andualem Mengistu.** 2019. "Effective Tax Rates and Firm Size in Ethiopia." *Development Policy Review* 37 (S2): O248–73.
- Mascagni, Giulia, Andualem T. Mengistu, and Firew B. Woldeyes.** 2021. "Can ICTs Increase Tax Compliance? Evidence on Taxpayer Responses to Technological Innovation in Ethiopia." *Journal of Economic Behavior and Organization* 189: 172–93.
- Mayega, Jova, Ronald Waiswa, Jane Nabuyondo, and Milly Nalukwago Isingoma.** 2021. "How Clean Are Our Taxpayer Returns? Data Management in Uganda Revenue Authority." International Centre for Tax Development African Tax Administration Paper 24.
- McKenzie, David, and Anna Luisa Paffhausen.** 2019. "Small Firm Death in Developing Countries." *Review of Economics and Statistics* 101 (4): 645–57.
- Mitchell, Brian R.** 2007. *International Historical Statistics*. London: Palgrave Macmillan.
- Moore, Mick, Wilson Prichard, and Odd-Helge Fjeldstad.** 2018. *Taxing Africa: Coercion, Reform and Development*. London: Bloomsbury Academic.
- Naritomi, Joana.** 2019. "Consumers as Tax Auditors." *American Economic Review* 109 (9): 3031–72.
- OECD.** 2023. "Inventory of Tax Technology Initiatives." <https://www.oecd.org/tax/forum-on-tax-administration/tax-technology-tools-and-digital-solutions/> (accessed on April 27, 2023).
- Okunogbe, Oyebola.** 2023. "Becoming Legible to the State: The Role of Identification and Collection Capacity in Taxation." World Bank Policy Research Working Paper 9852.
- Okunogbe, Oyebola, and Victor Pouliquen.** 2022. "Technology, Taxation, and Corruption: Evidence from the Introduction of Electronic Tax Filing." *American Economic Journal: Economic Policy* 14 (1): 341–72.
- Okunogbe, Oyebola, and Fabrizio Santoro.** 2023a. "The Promise and Limitations of Information Technology for Tax Mobilization." *World Bank Research Observer* 38 (2): 295–324.
- Okunogbe, Oyebola, and Fabrizio Santoro.** 2023b. "Increasing Tax Collection in African Countries: The Role of Information Technology." *Journal of African Economics* 32 (S1): i57–83.
- Oyebola Okunogbe, and Gabriel Tourek.** 2023. "Reproducibility package for How Can Lower-Income Countries Collect More Taxes? The Role of Technology, Tax Agents, and Politics." World Bank Reproducible Research Repository. <https://doi.org/10.60572/c73wj635>.
- Olawoyin, Oladeinde.** 2020. "Tinubu, Alpha-Beta, Others Accused of Fraud, Tax Evasion, Money Laundering." *Premium Times*, October 8. <https://www.premiumtimesng.com/news/>

- headlines/419179-tinubu-alpha-beta-others-accused-of-fraud-tax-evasion-money-laundering.html (accessed July 14, 2023).
- Prichard, Wilson.** 2018. "Electoral Competitiveness, Tax Bargaining and Political Incentives in Developing Countries: Evidence from Political Budget Cycles Affecting Taxation." *British Journal of Political Science* 48 (2): 427–57.
- Prichard, Wilson.** 2019. "Tax, Politics, and the Social Contract in Africa." In Oxford Research Encyclopedia of Politics. <https://oxfordre.com/politics/display/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-853>.
- Regan, Tanner and Priya Manwaring.** 2023. "Public Disclosure and Tax Compliance: Evidence from Uganda." Institute for International Economic Policy Working Paper 2023-04.
- Santoro, Fabrizio, Adrienne Lees, Marco Carreras, Theonille Mukamana, Naphtal Hakizimana, and Yves Nsengiyumva.** 2023. "Technology and Tax: Adoption and Impacts of E-services in Rwanda." ICTD Working Paper 153. Brighton: Institute of Development Studies. <https://doi/10.19088/ICTD.2023.001>.
- Shah, Jawad.** 2023. "Using Computerized Information to Enforce VAT: Evidence from Pakistan." Unpublished.
- Suri, Tavneet, and William Jack.** 2016. "The Long-Run Poverty and Gender Impacts of Mobile Money." *Science* 354 (6317): 1288–92.
- UNU-WIDER.** 2022a. "Government Revenue Dataset." Version 2022. <https://doi.org/10.35188/UNU-WIDER/GRD-2022> (accessed on October 14, 2023)..
- UNU-WIDER.** 2022b. "World Income Inequality Database (WIID)." <https://doi.org/10.35188/UNU-WIDER/WIID-300622> (accessed June 30, 2022).
- Weigel, Jonathan L.** 2020. "The Participation Dividend of Taxation: How Citizens in Congo Engage More with the State When It Tries to Tax Them." *Quarterly Journal of Economics* 135 (4): 1849–1903.
- Wier, Ludvig S., and Gabriel Zucman.** 2022. "Global Profit Shifting, 1975–2019." NBER Working Paper 30637.
- World Bank.** "Enterprise Surveys." <http://www.enterprisesurveys.org> (accessed on June 2, 2023)..
- World Bank.** 2023. "World Development Indicators." Version 2023. Washington, DC: World Bank. <https://databank.worldbank.org/source/world-development-indicators> (accessed on May 1, 2023).

# Does the Value-Added Tax Add Value? Lessons Using Administrative Data from a Diverse Set of Countries

Anne Brockmeyer  Giulia Mascagni  Vedanth  
Nair  Mazhar Waseem  Miguel Almunia

**T**he value-added tax (VAT) has become a crucial source of revenue for most governments around the world, especially in lower-income countries: 175 jurisdictions have adopted a VAT as of 2023, up from just 30 at the beginning of the 1980s (Caragher 2023; Ebrill et al. 2001). In 2019, the VAT raised 37 percent of total tax revenue (including social contributions) on average in low-income countries and 22 percent in high-income countries (UNU-WIDER 2023). International organizations such as the International Monetary Fund (IMF) and the World Bank have played a key role in advising countries to adopt the VAT (Baunsgaard and Keen 2010; Brautigam, Fjeldstad, and Moore 2008) to increase domestic resource mobilization and replace tariff revenues lost in the process of trade liberalization. The only major economy without a VAT is the United States.

The widespread adoption of the value-added tax globally has been justified by its appeal as a tax that is both more efficient (because it does not distort firms’

■ *Anne Brockmeyer is the Global Lead for Tax Data Analytics at the World Bank, Research Fellow at the Institute for Fiscal Studies and Honorary Associate Professor, University College London. This work was conducted while she was Research Director at the Centre for Tax Analysis in Developing Countries, Institute for Fiscal Studies (IFS), London, United Kingdom. Giulia Mascagni is Research Director, International Centre for Tax and Development (ICTD) and Research Fellow, Institute of Development Studies, both in Brighton, United Kingdom. Vedanth Nair is a Research Economist at the Centre for Tax Analysis in Developing Countries, Institute for Fiscal Studies (IFS), London, United Kingdom. Mazhar Waseem is Associate Professor of Economics, University of Manchester, Manchester, United Kingdom. Miguel Almunia is Associate Professor of Economics, CUNEF Universidad, Madrid, Spain. Their email addresses are [abrockmeyer@worldbank.org](mailto:abrockmeyer@worldbank.org), [g.mascagni@ids.ac.uk](mailto:g.mascagni@ids.ac.uk), [vedanth\\_n@ifs.org.uk](mailto:vedanth_n@ifs.org.uk), [mazhar.waseem@manchester.ac.uk](mailto:mazhar.waseem@manchester.ac.uk), and [miguel.almunia@cunef.edu](mailto:miguel.almunia@cunef.edu)*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.1.107>.

production choices) and easier to enforce (due to the data trails it generates) than other indirect taxes. However, the desirable properties of a “textbook” VAT may be diminished in the context of low- and middle-income countries (henceforth, “lower-income countries”). These economies feature widespread business and employment informality, limited administrative capacity of both firms and the government, and liquidity constraints. Although basic public finance models are often based on the assumption that taxes are perfectly enforced at no cost, this assumption is unreasonable in most contexts—and especially so in lower-income countries. Real-world VAT systems may also depart from the textbook model due to policy choices, such as the introduction of multiple VAT rates (with the aim of making the tax more progressive or providing tax relief to specific groups of firms), along with limits to firms’ ability to obtain tax refunds when they have negative VAT liabilities. It is therefore important to assess how the VAT is performing in practice, and how its performance in lower-income countries compares to higher-income contexts.

As value-added taxes spread across the developing world, an emerging and influential literature has sought to examine their real-world effects (Bird and Gendron 2007; Keen 2007; Ebrill et al. 2001; Tait 1998). In the past, this research typically relied on aggregate data, which meant that it could not identify how individuals or firms respond to the tax system. As administrative data—which are (usually confidential) data collected as part of the tax collection process—became increasingly available to researchers focusing on lower-income countries, a recent literature has emerged to use these data to document departures of the real-world VAT from the textbook model.<sup>1</sup> This paper builds upon the existing literature by providing systematic evidence on the discrepancies between the textbook model of a VAT and its real-world implementation, leveraging micro-level administrative data from VAT records in eleven countries at different income levels, ranging from Ethiopia, with a GDP per capita of \$500, to France, with a GDP per capita of \$45,000. This is, to our knowledge, the first paper that uses VAT administrative data from a wide range of countries and is part of a larger agenda using cross-country administrative data to generate novel facts on public finances and firms (for example, Bachas et al. 2023; Bachas, Brockmeyer, and Semelet 2020).

We start by describing how the value-added tax works and providing a precise definition of what we call the “textbook” VAT model. We then present four empirical facts on the real-world implementation of the VAT, based on the administrative microdata. Finally, we discuss the results of counterfactual policies that involve replacing the VAT with alternative tax instruments, like a retail sales tax or a turnover tax, and outline policy implications and avenues for further research. Despite its shortcomings in the context of lower-income countries, we conclude that the real-world VAT is superior to the alternatives.

<sup>1</sup>A nonexhaustive list of recent studies in this literature includes Waseem (2023), Carrillo et al. (forthcoming), Mascagni et al. (2022), Liu et al. (2021), Benzarti et al. (2020), Almunia et al. (2022), Gadenne, Nandi, and Rathelot (2022), Gerard et al. (2022), and Agrawal and Zimmermann (2019).

## Value-Added Tax 101: A Primer

### How Does a Value-Added Tax Work?

A value-added tax seeks to tax the value added at each stage of the production chain where it is generated. Specifically, firms use labor and intermediate inputs to produce outputs they sell, either to other firms or to final consumers. The difference between the value of a firm's output and that of its intermediate nonlabor inputs constitutes its value added. In a closed economy, if all value added is taxed at each step in the production chain, the result is equivalent to a sales tax on final consumption imposed at the retail stage because, in an accounting sense, the value of the final consumption good is the same as the sum of value added at each stage of the production chain. Thus, the VAT is often referred to as a tax on consumption.

In the basic textbook version, the value-added tax is applied to all transactions in the economy. When firms buy and sell, the invoices specify both a price and the VAT that is being charged. Each firm has to submit a monthly (sometimes quarterly or annual) VAT declaration where it reports the VAT that it collected when selling its output ("output VAT") and the VAT it paid when purchasing inputs from suppliers ("input VAT"). In a standard VAT system, firms can credit the input VAT they paid to suppliers against the output VAT they collected when selling to customers. Thus, firms only remit to the government the difference between output and input VAT ("net VAT"), ensuring that the tax applies only to a firm's value added and not to intermediate inputs that were already taxed in previous stages in the production chain. Crawford, Keen, and Smith (2010) provide a more detailed outline of the textbook value-added tax.<sup>2</sup>

One challenge for the value-added tax is that business activities and transactions often cross national borders. Thus, policymakers must determine how imports and exports would be treated by their national tax systems. Because the VAT is intended as a tax on final consumption, most VAT systems follow the so-called "destination principle," by which the tax liability on a transaction is attributed to the jurisdiction in which consumption occurs. Accordingly, countries apply VAT on imports, as imports will be consumed in that country's domestic market (either by final consumers or as intermediate inputs), but they do not normally apply VAT on exports because they will be consumed in the destination country that will tax them according to their own laws. Therefore, we say that exports are "zero-rated" under the VAT. As a consequence of the destination principle, exporting firms often end up with a negative VAT liability: they do not collect output VAT on their export sales, but they pay VAT on their inputs (both domestic and imported). A negative VAT liability may also arise for firms that make losses or large capital purchases in a period. In these cases, firms should be able to request a VAT refund from the government.

<sup>2</sup>Financial services firms are usually exempt from real-world value-added taxes or are subject to special rules, as are governments, charities, and research organizations. For a discussion of the merits of these exemptions, see Mirrlees et al. (2011).

### Why Is the Value-Added Tax Attractive?

The textbook value-added tax satisfies two concepts of efficiency: production efficiency and revenue efficiency.

Production efficiency means that the economy is at its production frontier, so that there is no way to increase the production of one good without decreasing the production of another good. For a tax to satisfy this property, it must not distort firms' production decisions: the tax should not favor the use of one type of input over another (say, domestic versus imported), nor the production by certain firms over others, nor vertical integration over a distributed supply chain (Diamond and Mirrlees 1971). A value-added tax is production efficient because it does not tax intermediate inputs and, as a result, all firms face the same relative prices and will choose the same input mix regardless of the tax. The ability of firms to credit input VAT against output VAT, so that they only remit net VAT to the government, is key to this production efficiency. In contrast, a turnover tax that applies to all sales (including intermediate inputs) would create incentives for vertical integration—a distortion that would make it production inefficient—because taxes would keep accumulating at each stage of production, resulting in a greater tax burden on consumption (the final stage) for production chains that have more intermediate stages upstream. This phenomenon is called “tax cascading.”

Revenue efficiency relates to whether a tax is robust to evasion or, conversely, how well it is able to generate the maximum amount of revenue for a given amount of administrative effort (Best et al. 2015). In the design of the value-added tax, all transactions between two firms should be reported to the government twice—once by the seller and once by the buyer. The government can cross-check the two reports in order to detect potential misreporting. Government use of these so-called “third-party” reports—paper trails created by agents other than the taxpayer in question—has been a prominent subject of the recent tax enforcement and development literature (for example, Naritomi 2019; Pomeranz 2015; Brockmeyer et al. 2019). Moreover, because a VAT generates a large amount of data along the production chain, it lends itself particularly well to technological methods of facilitating data-based enforcement. For instance, e-invoicing and electronic billing machines record transactions, allowing the government to cross-check reports with relative ease and target tax enforcement accordingly.<sup>3</sup>

The paper trails generated by a value-added tax are particularly interesting from a tax enforcement perspective, as the buyer and seller in a transaction have asymmetric incentives to misreport. The buyer would like to overreport the transaction amount to reduce their tax liability, while the seller would like to underreport the transaction amount for the same reason. These asymmetric incentives should

<sup>3</sup>E-billing does seem to improve tax compliance, especially when coupled with more traditional forms of verification (Mascagni, Mengistu, and Woldeyes 2021). However, third-party reporting is not a panacea for improved tax enforcement if taxpayers can misreport on margins not reported by third parties (Carrillo, Pomeranz, and Singhal 2017; Slemrod et al. 2017), if third-party reporting covers only a subset of transactions (Brockmeyer and Sáenz Somarriba 2022), or if the revenue authority lacks the administrative capacity to do systematic cross-checks (Almunia et al. 2022).



prevent a buyer and seller from colluding to misreport the transaction.<sup>4</sup> In addition, the VAT remitted at upstream stages acts as a withholding tax for registered downstream firms, giving them an incentive to report output VAT at least as large as input VAT. Compliance should thus propagate down the supply chain (Waseem 2022). The revenue efficiency of a VAT is aided by the fact that it is levied in small chunks along the entire supply chain, so a substantial drop in VAT revenue will only occur if many firms along the supply chain are noncompliant.

### **VAT in the Real World: Four Facts**

We have already referred to the “textbook” model of a value-added tax. Such a model is characterized by the following features: (1) universal coverage (no exemptions for goods, for smaller firms, or for specific industries); (2) a uniform tax rate; and (3) automatic, costless refund of negative tax liability. A VAT with these features would be neutral to production and consumption decisions in the economy.

However, most real-world value-added tax systems have features that move away from the textbook model. Firms with sales below a certain threshold, for example, are usually not required to register for the VAT; some goods may be exempt from the tax; others may be taxed at a rate below the standard VAT rate; firms may fail to claim input VAT; and refunds may take time to process. In addition to these differences, a substantial fraction of firms in lower-income countries are informal—that is, not registered with the tax administration—and hence do not submit any tax declarations. Overall, taxpayers likely differ in their knowledge about the tax system and in the financial, organizational, and cognitive resources they can draw on to comply with taxes.

In this section, we document four facts that highlight some of the discrepancies between textbook and real-world value-added tax systems. Our analysis is based on administrative VAT-returns data from eleven countries across a range of per capita incomes: Costa Rica, Eswatini (formerly Swaziland), Ethiopia, France, Guatemala, Honduras, Hungary, Pakistan, Rwanda, Senegal, and Uganda. The data should cover all VAT-filing firms in those countries, but typically do not cover VAT collected at the import stage. At the minimum, the declarations include information on a firm’s total sales, output VAT, and input VAT. Unless otherwise specified, we will focus on “net” VAT, equal to output VAT minus input VAT, as our measure of tax liability, because this measure is the most comparable across countries. The presence of other variables—such as total purchases, the breakdown of sales into various categories, and credits carried to or from other periods—is not universally available across countries, so not all results are available for every country in our sample. For Pakistan, Rwanda, and Uganda, in addition to declaration-level data, we have access to transaction-level data, which include every sale and purchase reported by a firm

<sup>4</sup>Note that these incentives break down at the final consumption stage. In this case, final consumers cannot credit the value-added tax paid against their own tax liability, so there is more room for collusion between sellers and buyers not to report the transaction for VAT purposes.

to and from other VAT-registered firms. We typically have five years of data for each country, but the available years do not overlap perfectly across countries. For results in which we present a single year for each country, we take the latest available year prior to 2020 for each country, to avoid contamination by the Covid-19 pandemic. The online Appendix for this paper discusses the choices we made in preparing the data and presents additional analyses and robustness tests for many of the facts presented in this paper.

### **Fact # 1: VAT Revenue Is Highly Concentrated on the Largest Taxpayers**

The value-added tax is designed to be broad based, remitted by each firm in the economy in proportion to its value added. The duty of remitting the VAT is meant to be distributed across many firms, in contrast to a retail sales tax, which imposes the burden of remitting the tax on the retail sector alone. Spreading the burden to all firms is thought to protect the VAT against evasion and allow it to raise a substantial amount of revenue. However, in most of the countries included in our analysis, more than 90 percent of VAT revenues are remitted by the largest 10 percent of firms, and this pattern is especially strong in lower-income countries.

Figure 1, panel A, plots the share of value-added tax revenue contributed by the largest 10 percent of firms (by total VAT remitted) against GDP per capita (measured on a log scale). Each dot represents a country-year observation, and there are multiple observations for each country. In lower-income countries such as Ethiopia or Uganda, the largest 10 percent of taxpayers account for 90–95 percent of VAT revenue, and for Pakistan this share reaches up to 99 percent.<sup>5</sup> In high-income countries such as France and Hungary, the level of concentration is somewhat smaller: the largest 10 percent of taxpayers account for around 85 percent of total VAT revenue.<sup>6</sup> Because revenue from the VAT is concentrated in a small number of firms, it may be highly sensitive to changes in the growth rate (or tax compliance behavior) among the top tax remitters.

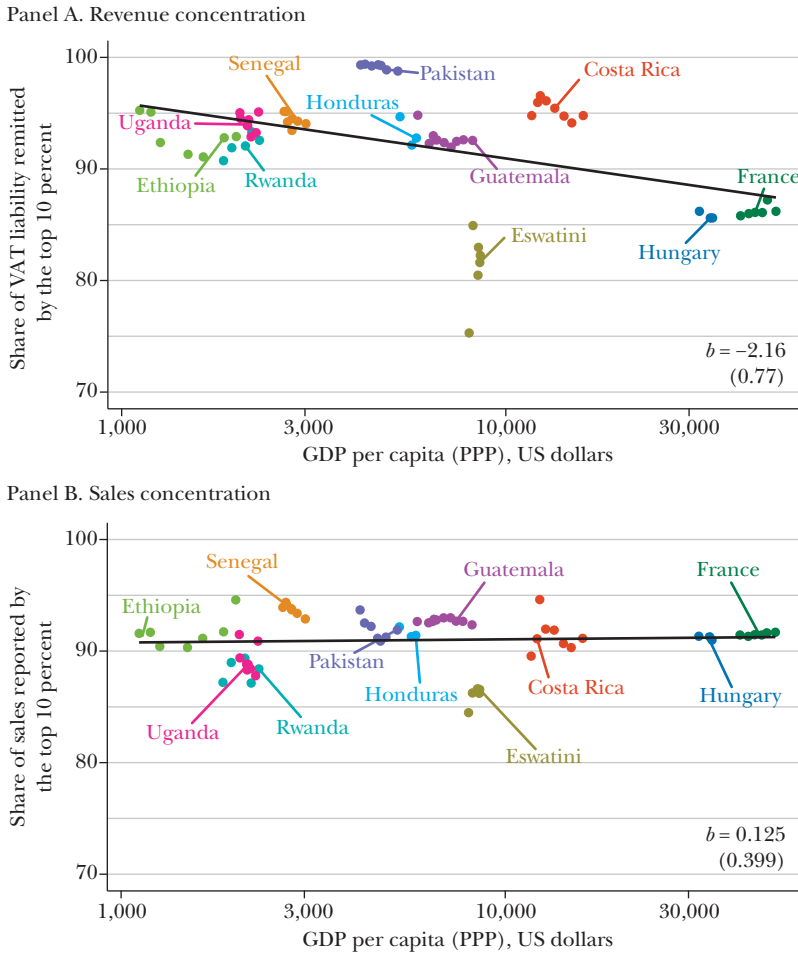
A natural explanation for the concentration of value-added tax revenue is that the firm size distribution is also very concentrated. Figure 1, panel B, shows that the largest 10 percent of firms report about 90 percent of sales revenue in all countries in our sample. While the concentration of sales is indeed very high, it is not correlated with GDP per capita. Thus, features of VAT design and implementation may be more important than economic structure in explaining the differences in VAT revenue concentration between high- and lower-income countries.

What explains the stronger concentration of value-added tax revenue in lower-income countries? One potential explanation is the different levels of administrative capacity across countries. When such capacity is limited, it is rational for governments to focus most enforcement efforts on the largest taxpayers, since the expected

<sup>5</sup>Eswatini (former Swaziland), a small country in southern Africa, features a lower degree of VAT revenue concentration and it is an outlier in this dimension.

<sup>6</sup>Using publicly available information, we observe that the corresponding share in 2019 was 84 percent in the United Kingdom (HMRC 2018–2019) and 87 percent in Spain (AEAT 2019). Thus, France and Hungary appear to be representative of other high-income European countries, rather than outliers.

Figure 1  
**Revenue and Sales Concentration**



Source: Derived from VAT declaration data from each country.

Note: Panel A plots the share of a country’s VAT liability that is contributed by the largest 10 percent of firms against country GDP per capita (on a log scale). Each dot represents a country-year observation. Shares are calculated using firm-level data aggregated by calendar year, for firms with positive sales. The VAT liability is defined as  $\max(0, \text{output VAT} - \text{input VAT})$ . This is our preferred measure of tax liability, as it is not affected by withholding systems and credits being carried forward from previous periods, and is therefore the most comparable across countries. In addition, this method allows us to calculate revenue concentration without drawing on refund data, which is not available for all countries. The largest 10 percent of firms are those that have the largest tax liabilities. Panel B plots the share of reported sales that are contributed by the largest 10 percent of firms, for firms with positive sales. The largest 10 percent of firms are those that have the largest sales. The slope coefficient displayed on the graph is from a simple regression of  $y$  on  $x$ , with standard errors clustered at the country level.

return of auditing them is higher in terms of additional tax revenue. Indeed, while revenue authorities in all countries spend relatively more resources on the enforcement of large taxpayers compared to small ones, the focus on the former is stronger for lower-income countries (Bachas, Fattal Jaef, and Jensen 2019). However, it is difficult to quantify the importance of this factor, because administrative capacity is hard to measure in a consistent manner across countries.

A second potential explanation is the existence of registration thresholds for the value-added tax. Most countries establish a minimum size threshold below which firms are not required to register in the VAT system. Firms below the threshold are allowed to register voluntarily for the VAT or, in some cases, can opt into a simplified tax where the tax base is total sales revenue. Assuming that the revenue collected from small firms is negligible, having a lower registration threshold would mechanically increase the share of revenue from the largest 10 percent of firms. In order to compare VAT registration thresholds across countries, we divide the level of their thresholds by each country's GDP per capita. According to this metric, registration thresholds are significantly higher in lower-income countries, with a median threshold of 1,600 percent of GDP per capita, in contrast to 50 percent of GDP per capita amongst high-income countries. Thus, the level of VAT registration thresholds is unlikely to drive the negative correlation between VAT revenue concentration and GDP per capita.

The exclusion of small firms from the value-added tax is a clear departure from the textbook model and distorts production efficiency, as VAT-registered firms remit tax on sales to unregistered firms, but the latter cannot claim back any tax paid on their inputs. Unregistered firms are therefore incentivized to purchase intermediate inputs from other unregistered firms or to use more labor inputs, which are untaxed under the VAT. There is indeed evidence that the VAT does provide incentives for registered and unregistered firms to operate in separate supply chains, as de Paula and Scheinkman (2010) show empirically in Brazil and Gadenne, Nandi, and Rathelot (2022) show in India. Relatively higher VAT registration thresholds in lower-income countries likely increase the magnitude of supply chain segmentation and further diminish production efficiency.

If setting a registration threshold for the value-added tax hurts production efficiency, why is this policy so common? The main reason is that the cost of complying with the VAT is disproportionately higher for small firms than larger ones (Coolidge 2012; Yesegat et al. 2016; Highfield et al. 2019). Moreover, the administrative cost of managing a VAT is up to three times larger in low-income countries than in high-income ones (Crandall, Gavin, and Masters 2021). Thus, setting the VAT registration threshold at the optimal level requires trading off the production inefficiency and revenue loss generated by exempting small firms from the VAT system against the compliance and administrative costs created by including them (Keen and Mintz 2004). These costs are absent in the textbook VAT model, but important in the real world.

In sum, the concentration of revenues from the value-added tax is extremely high among all countries, and particularly so in lower-income countries. These results suggest that the burden of the VAT is not spread as broadly across firms as

one might think. A potential reason for this pattern that we have not yet considered is that the tax applies broadly to (almost) all transactions of firms registered for this tax. As we discuss in the next section, however, in many countries a substantial share of transactions among registered firms are exempt or subject to a reduced VAT rate.

**Fact # 2: Effective Tax Rates Are Lower for Larger Firms**

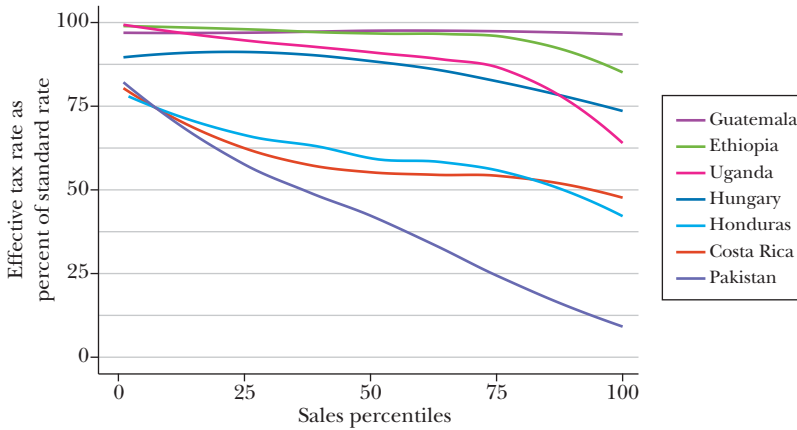
A key feature of the textbook value-added tax is that all transactions of goods and services are covered by the same tax rate, called the “standard” VAT rate. Exemptions for certain goods or services move the VAT away from a broad tax on all value added and distort input choices of firms. In addition, multiple VAT rates distort consumers’ final consumption choice—by inducing them to consume relatively more low-rated goods—and thereby also affect firms through the demand channel.<sup>7</sup> There is some evidence that eliminating VAT exemptions and reduced rates and redistributing the resulting revenue through a means-tested transfer program would increase welfare (Crawford, Keen, and Smith 2010). Yet VAT exemptions and reduced rates are found in abundance in almost all countries. Governments justify them on multiple grounds, such as to reduce the tax burden on the poor (exemption or reduced rating of food and utilities) or to promote the consumption of goods and services they perceive will benefit the economy (exemption or reduced rating of books and information technology services).

To assess the extent to which real-world value-added tax systems depart from the principle of uniform taxation, we examine the difference between standard and “effective” VAT rates. We calculate the effective tax rate as the actual net VAT collected, divided by the total value added. More specifically, using our firm-level data, we can look at dispersion in effective tax rates across firms of different sizes and across industries. We then discuss the efficiency consequences of the observed pattern of effective tax rates.

Effective value-added tax rates can vary due to rate differentiation or exemptions across goods and services. The variation is hence driven by a policy choice to offer exemptions and reduced rates and by firms’ decisions to report selling exempt and reduced-rated goods, which may or may not correspond to their actual sales. Our measure of the effective tax rate should not be affected by evasion, assuming that a firm seeking to underreport its net value-added tax would also proportionally reduce its value added, so as to not attract the attention of the tax authorities (though it is possible that firms make mistakes). In our calculations of effective tax rates, we exclude exporters (defined as firms with exports worth at least 30 percent of their sales), because exports are always zero-rated under the destination-based principle, so zero-rating of exports cannot be thought of as a deviation from the textbook VAT.

<sup>7</sup>Rate differentiation for a consumption tax may be optimal from a welfare-maximizing perspective if a linear income tax is the only available tool aside from the consumption tax (Ramsey 1927). However, Atkinson and Stiglitz (1976) show that a uniform consumption tax rate is optimal if the policymaker can also use a nonlinear income tax and if the consumers’ utility function is weakly separable in goods and leisure.

Figure 2

**Effective Tax Rates as a Percent of Standard Tax Rate**

Source: Derived from VAT declaration data from each country.

Note: This figure shows effective tax rates as a percentage of the statutory tax rate (STR), by firm size (total sales) percentiles, fitted with a LOESS (“locally estimated scatterplot smoothing”) polynomial. For the actual values of effective tax rate by firm size percentile and the fitted curve, see the series “actual” in online Appendix Figure A.4. Effective tax rate is defined as annual net VAT over annual value-added, where net VAT = (output VAT - input VAT), and value added = (total reported sales - total reported purchases). Effective tax rate are winsorized at the standard rate (or at the higher rate of VAT in Honduras). Exporters (defined as those where annual exports are >30 percent of total sales) are excluded, as the zero-rating of exports is taken to be part of the “textbook” VAT system. Including exports does not substantially alter the pattern observed. Note that input nonclaiming—as described in fact #3—does not affect this result, as nonclaiming on inputs is also associated with nonclaiming of purchases, so net VAT is in line with value added. The statistics shown in this figure are not available for France, Senegal, or Rwanda as VAT declaration data in these countries do not include total purchases, so value added cannot be calculated. Results for Eswatini are not included due to issues with large numbers of firms paying ETRs greater than the standard rate.

Figure 2 shows that the effective value-added tax rate is below the statutory rate in most of the countries in our sample, and that the gap decreases with firm size—although with substantial variation across countries. In Honduras, firms at the 10th percentile of firm size face an effective tax rate that is 75 percent of the standard rate, compared to firms at the 90th percentile, who face an effective tax rate that is 45 percent of the standard rate. In Ethiopia, firms at the 10th percentile face an effective tax rate almost exactly equal to the standard rate, while firms at the 90th percentile face an effective tax rate that is 90 percent of the standard rate. These patterns stand in contrast to effective tax rates for the corporate income tax, which Bachas et al. (2023) found to follow an inverse U-shape, with higher effective tax rate for firms in the middle of the distribution and lower effective tax rate for small firms and very large firms.

The effective value-added tax rate can be lower for larger firms either because larger firms are more likely to sell products that happen to be exempt or reduced-rated or because smaller firms are less likely to report selling exempt or reduced-rated goods. The reporting explanation seems more likely, because a closer look at this

data shows that even in narrowly defined sectors where large firms are likely to produce the same goods as small firms, larger firms have much lower effective tax rates than small firms. Small firms may not claim exemption or reduced-rate (even if the goods they sell are eligible) either because they lack the knowledge that such provisions exist or because the compliance costs associated with keeping track of sales of exempt, reduced-rated, and standard-rated goods are too large. In the presence of high compliance costs, small firms may be willing to trade off a higher tax liability for reporting simplicity.

In most countries in our sample, exemptions are more important (as a percentage of total sales) than differentiated rates in driving a wedge between the effective and the standard value-added tax rates. Exceptions to this pattern are high-income France and Hungary and middle-income Eswatini, which rely more on reduced rates. VAT exemptions have long been recognized as a source of substantial departure of real-world VATs from the textbook model, especially if they occur in the middle of the production chain (Ebrill et al. 2001). Consider the case of a farmer that sells wheat to a mill, which then sells flour to a bakery, which sells bread to consumers. If flour (but not wheat) is exempt from VAT, then the farmer remits VAT on sales to the mill but the mill cannot reclaim the input VAT. The mill does not remit VAT when selling to the bakery, and the bakery remits VAT when selling to the final consumer but, again, cannot reclaim any input VAT. In effect, the value added by the farmer has been taxed twice—when selling to the mill, and again when the bakery sells to the final consumer without reclaiming input VAT. A vertically integrated firm, which does farming, refining, and baking itself, faces a lower tax liability than a production chain with three separate firms.<sup>8</sup> This example also highlights how seemingly well-meaning political rallying cries such as “don’t tax the poor” or “don’t tax small farmers” may end up hurting the sectors they seek to protect.

In addition to distorting firms’ production choices, exemptions and reduced rates also affect the revenue efficiency of the value-added tax in three ways. First, they result in a loss of revenue due to a lower effective tax rate. In lower-income countries, exemptions and reduced rates cost on average 30 percent of total VAT revenue, though with substantial variation between countries of a similar level of income. Second, as discussed earlier, the revenue efficiency of the VAT is, in theory, generated by interlinkages between firms. Exemptions and reduced rates can create breaks in the VAT chain, allowing firms that are upstream from exempt firms the opportunity to evade. Third, exemptions and reduced rates increase the complexity of the VAT system, which at a minimum increases the administrative burden of the tax and at worst creates an opportunity for evasion by product misclassification (Fisman and Wei 2004). In one case, the UK tax authority had to clarify that children’s clothing made from goat fur was zero-rated, with the exception of goat fur from Mongolia, Tibet, or Yemen, which was standard-rated.

Equity is also a relevant issue. A growing literature has studied the effect of value-added tax exemptions and reduced rates on consumers. For example,

<sup>8</sup>For a deeper discussion of this example, see Iddrisu, Parekh, and Phillips (2023).

widespread VAT exemptions for food are regressive in lower-income countries, as poorer consumers purchase food from the informal sector, which does not directly benefit from VAT exemptions (as discussed in this issue by Bachas, Jensen, and Gadenne). There has been less focus on equity between firms, though this margin may be particularly relevant in lower-income countries, where larger firms tend to have richer owners, hire more formal workers, and pay higher wages (Ulyssea 2018; Brown and Medoff 1989). The fact that larger firms seem to benefit the most from VAT exemptions and rate differentiation weakens the rationale for using the VAT as a tool for redistribution.

In short, exemptions and reduced rates create a gap between the effective and standard value-added tax rates. The existence of this gap, and the fact that its size is strongly correlated with firm size, imply that the VAT distorts firms' production choices, is less revenue-efficient and can make the VAT less equitable.

### **Fact # 3: Nonclaiming of Input VAT Is Common among Small Firms**

With a broad-based value-added tax, all firms (except those using only labor inputs) should claim input VAT. This feature is important for production efficiency because accounting for input VAT is what ensures that the tax does not distort firms' input mix nor create incentives for vertical integration. The feature also matters for revenue efficiency because the practice of reporting each transaction twice, by the buyer and the seller in a transaction, generates information trails used for tax enforcement.

However, a substantial share of firms do not claim input VAT, despite reporting positive sales. Small firms are especially likely to avoid doing so, as shown in panel A of Figure 3. Across countries, around 40 percent of the smallest firms in the VAT (those in the bottom 5 percent of the sales distribution) do not claim any input VAT. The pattern is particularly striking in some countries, such as Senegal, where close to 80 percent of small firms have no input claim. Input nonclaiming is generally more common in lower-income countries than in higher-income countries, as shown in panel B.<sup>9</sup>

Firms may choose not to claim input value-added tax for several reasons. A firm should not claim input VAT if it did not make any input purchases or if it purchased only untaxed goods, such as goods supplied by informal or unregistered firms, or goods that are exempt from VAT. Indeed, the nonclaiming of input VAT by small firms can in some cases be seen as a corrective feature, rather than a bug, of the VAT system (Keen 2008). For a production process in which goods move from the informal to the formal sector, the VAT indirectly taxes the value added in the informal sector, because formal firms cannot reclaim input VAT on purchases from informal firms.

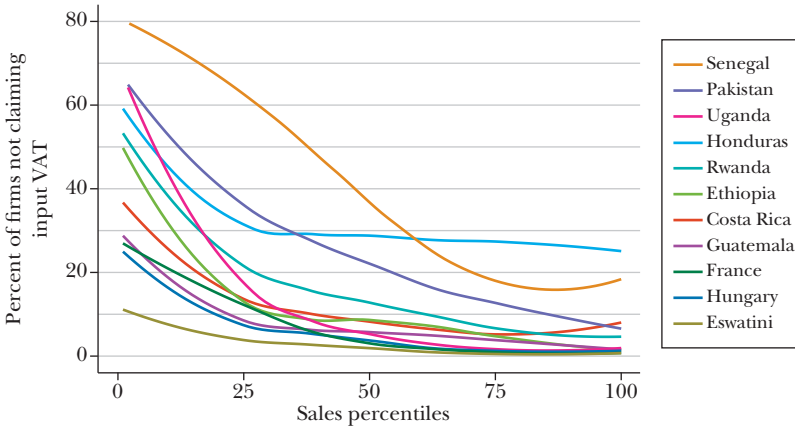
On the other hand, a situation in which firms do not claim input value-added tax for legitimate VAT-taxed purchases creates a clear departure from the textbook

<sup>9</sup>Note that input VAT nonclaiming does not drive the previous result on effective tax rates, as firms that do not claim VAT on inputs generally do not report any purchases, meaning that the effective tax rate, or the ratio of net VAT to value added, simply corresponds to the VAT rate applied on sales.

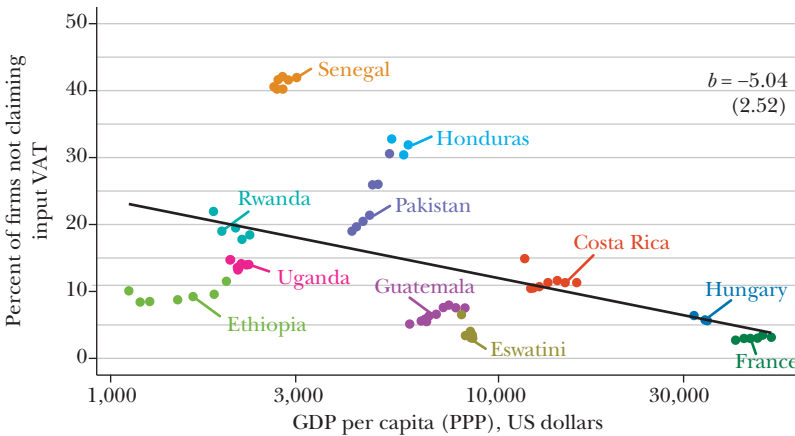


Figure 3  
**Input VAT Non-claiming**

Panel A. Within countries



Panel B. Across countries



Source: Derived from VAT declaration data from each country.

Note: Panel A plots the share of firms not claiming input VAT by firm size (total sales) percentiles, fitted with a LOESS (“locally estimated scatterplot smoothing”) local polynomial. For the actual values of input nonclaiming by percentiles and the fitted curve, see the series “baseline” in online Appendix Figure A.8. A firm has zero input VAT if they have declared zero input VAT over the entire year, conditional on reporting positive sales. Panel B plots the share of input VAT nonclaimers by country-year GDP per capita. The slope coefficient displayed on the graph is from a simple regression of  $y$  on  $x$ , with standard errors clustered at the country level.

VAT system. To see if firms are failing to claim input VAT paid on taxable transactions rather than simply not having any purchases or having exclusively nontaxable purchases, we use transaction-level data which detail all transactions that take place between VAT-registered firms. These data show that, of the firms which did not declare any input VAT on their VAT declaration, a large proportion—45 percent in Uganda (Almunia et al. 2022) and 79 percent in Rwanda (Mascagni et al. 2023)—could have claimed input VAT, as another VAT-registered firm did record selling to them. In a closer examination of this data, input nonclaiming is not associated with differences in production technology or in the likelihood of making nontaxable purchases across sectors. The correlation between input nonclaiming and firm size is almost unchanged when we control for the sectoral composition of firm-size groups.

Input nonclaiming is, in some ways, a paradox. Tax authorities typically assume that small firms underreport their tax liability, and tax evasion rates are indeed greater among small firms, as documented in Pakistan by Best, Shah, and Waseem (2022) and in Senegal by Bachas et al. (2022). However, by not claiming input value-added tax, some small firms seem to be “leaving money on the table.” As with the previous discussion, small firms may be trading off tax liability against compliance costs. Indeed, evidence from taxpayer interviews in Rwanda (Mascagni et al. 2023) suggests that firms do not claim input tax credit due to a lack of knowledge of the VAT system, failure to file receipts, inability to claim input VAT within the allowed time frame, and concerns that it will make the firm more likely to be audited.

What policy implications follow from this phenomenon? Input value-added tax nonclaiming breaks the production efficiency of the VAT. Intermediate inputs are effectively untaxed for large firms but are taxed for small firms that do not claim input tax credits. Goods produced by supply chains made up of small firms hence face a higher tax burden than goods produced by supply chains of larger firms. Input VAT nonclaiming also has equity implications, given that small firms tend to have poorer owners, workers, and customers. The impact on revenue efficiency is unclear. Input VAT reduces tax liability, so not claiming input VAT would increase tax revenue, all else equal. However, some firms may not be claiming input VAT because they are also underreporting their sales and they wish to appear small to the tax authority. Widespread input VAT nonclaiming, and broken VAT chains more generally, weaken the withholding mechanism of the VAT (Waseem 2022) and the ability of tax authorities to use third-party information from transaction-level data to monitor tax compliance.

These findings, along with the previous finding that small firms are less likely to claim exemptions or reduced rates, relate to a growing literature on suboptimal tax filing behavior. For example, in a study of small firms in Rwanda, Tourek (2022) finds that many report exactly the same amount of income tax each year. After a reform reduced the tax liability for small firms, a large share of treated firms continue to report the same amount of tax as they did previously. Our findings may also be the reason why many countries have established simplified tax regimes that allow small firms to pay an alternative tax, often a “turnover tax” based on total firm sales, instead of the value-added tax (Wei and Wen 2019). As noted earlier,

a turnover tax is a sales tax that is applied on sales of firms. It is usually applied at a rate lower than the VAT, but does not allow credit of tax paid on inputs and hence cascades through the supply chain. For instance, Ethiopia and Senegal have turnover tax regimes with rates between 2 and 10 percent, depending on the firm's sector. By not claiming input VAT, small firms are effectively paying a high-rate turnover tax, even when a more favorable alternative is available.

Being registered for the value-added tax may have advantages for small firms as well. As mentioned before, it provides them with access to a broader range of VAT-registered customers. In the context of the United Kingdom, Liu et al. (2021) show that, although being registered for VAT strictly increases tax liability for firms whose sales exceed their purchases, the general equilibrium benefits of VAT registration through increased sales to other firms outweigh the direct increase in tax liability. Consistent with this explanation, we find that a substantial share of firms in our data (30 percent in Pakistan, 65 percent in Senegal) appear to have voluntarily registered for the VAT despite having taxable sales below the relevant VAT registration threshold.

Input nonclaiming is a puzzling behavior. However, revenue authorities usually worry more about input overreporting, which can lead to illegitimate claims for refunds for the value-added tax. We turn to the issue of refunds in the next section.

#### **Fact # 4: Value-Added Tax Refunds Are Often Delayed**

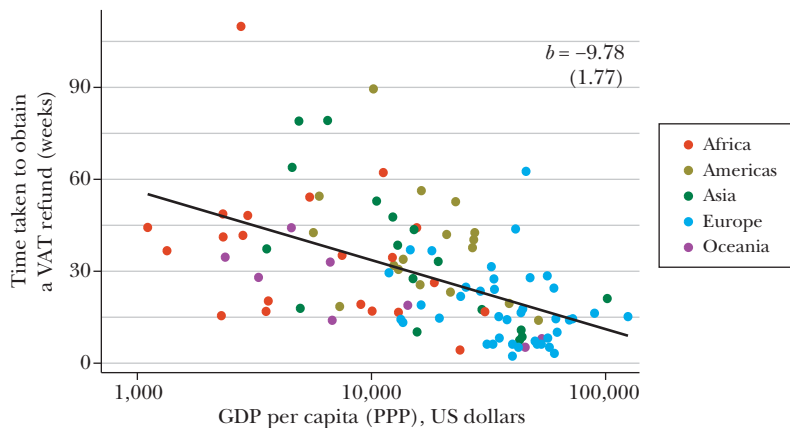
Refunds arise naturally in value-added tax systems, whenever a firm's input VAT credit exceeds its output VAT liability. This issue is often relevant for exporters due to the destination-based nature of the VAT. However, refunds can also arise for nonexporting firms if their purchases exceed sales in a given period or if they sell goods taxed at a reduced rate. (In addition, some VAT systems exempt large capital purchases from VAT to reduce the need to process refunds.) The textbook VAT model requires refunds to be processed quickly and costlessly, minimizing disruptions to firms' cash flow and production processes.

We draw on multiple data sources to examine the extent to which refunds arise in real-world value-added tax systems and how they are processed. We start by drawing on the World Bank Doing Business indicators, which rely on surveys of business leaders and accountants in 190 countries. Figure 4, panel A, shows that the time required for firms to receive a VAT refund is negatively correlated with the country's GDP per capita. A typical firm in a high-income country can expect a VAT refund within 10 weeks. In lower-income countries, the refund process takes an average of 45 weeks, with some extreme outliers, such as Pakistan, where the refund process takes 79 weeks on average.

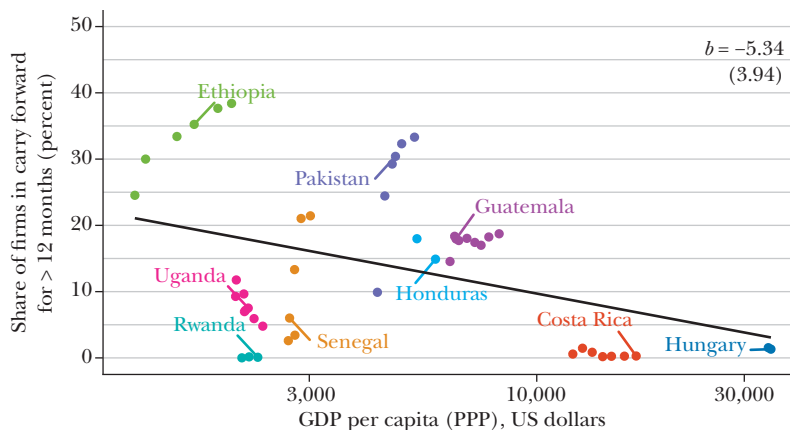
One explanation for refund delays is that governments need time to evaluate legitimate concerns about tax fraud that arise with a value-added tax. For most other taxes, evasion simply results in a low or zero tax liability. Under the VAT, tax evaders can instead claim large sums of money from the government in the form of refunds. Even high-income countries have struggled with this issue; for example, one study found that the United Kingdom and Germany were losing between 1 and 2 percent of VAT revenue due to "missing-trader" fraud (Keen 2007). In this form of fraud,

Figure 4  
Refund Delays

Panel A. World Bank Doing Business: Weeks taken to get a VAT refund



Panel B. Microdata: Percent firms carrying forward credits

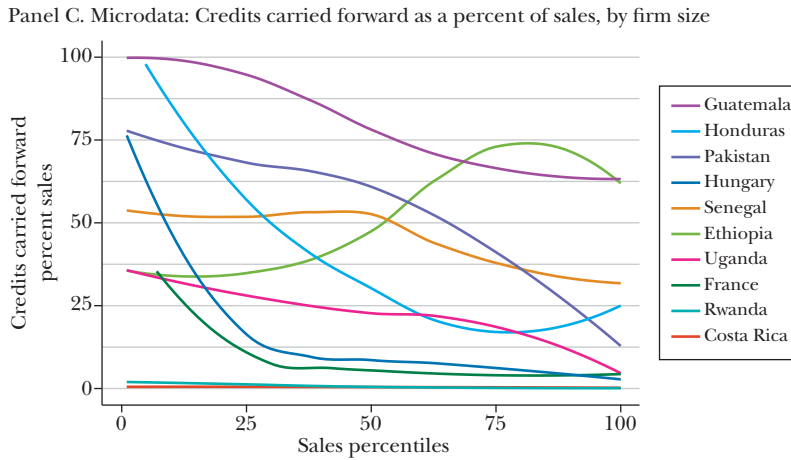


(continued)

a firm claims input tax credit for nonexistent “purchases” from a supplier firm, and then the firm disappears before remitting any output VAT. The challenge is even more severe in lower-income countries. Waseem (2023) finds that two-fifths of VAT refunds in Pakistan are based on invoices issued by “invoice mills”—fake firms that exist only to generate spurious value-added tax credits. Thus, tax authorities may respond to the threat of evasion through increased scrutiny of refund requests. In Rwanda and Uganda, for instance, all requests for value-added tax refunds trigger some type of verification.

Another widely held and more cynical view is that refund delays arise because cash-constrained governments seek to obtain free financing from firms. In most lower-income countries, tax authorities are not obliged to pay interest on delayed VAT refunds. Even in countries where tax authorities are required to pay interest on delayed VAT refunds (such as Indonesia, Kenya, and Zimbabwe), this does not

Figure 4  
Refund Delays (continued)



Source: Derived from VAT declaration data from each country.

Note: This figure documents VAT refund delays across and within countries. Panel A shows the number of weeks taken for firms to receive a VAT refund, based on the World Bank Doing Business Survey among business leaders and accountants (World Bank 2020; 2019 or latest year available). The findings are unchanged when we drop the four countries for which irregularities in the Doing Business indicators have been found. Panel B plots the share of firms (with positive sales) which carry forward VAT credits from previous periods in at least twelve consecutive VAT filings. Firms usually file VAT every month. We drop the small subsample of firms in specific countries that are allowed to file less frequently. The slope coefficient displayed on the graph is from a simple regression of  $y$  on  $x$ , with standard errors clustered at the country level. The statistics are not available for Eswatini due to data limitations. Panel C shows a LOESS (“locally estimated scatterplot smoothing”) local polynomial of the accumulated amount of credits carried forward over annual sales, by firm size, for the latest year of data available in each country. Online Appendix Figure A.12 reproduces this figure with the actual values of credits carried forward as percent of sales, by firm size percentiles, and the fitted line. In other graphs, the percentiles of sales for the  $x$ -axis are constructed by taking percentiles of average sales across the entire sample, due to concerns about measurement error in sales affecting both the  $x$  and  $y$  variable. Credits carried forward as a percent of sales is winsorized at 100 percent.

happen in practice (Pessoa et al. 2021). As the stock of refund requests can be as high as 30 percent of gross VAT receipts (Pessoa et al. 2021), with the funds already sitting in government coffers, it is a tempting pot of money for governments to tap into. As one example, Pessoa et al. (2021) claim that indebted countries such as Greece delayed VAT refunds after the 2008 financial crisis. In lower-income countries, where budgets are already heavily constrained, paying all refund claims might not even be feasible without reducing public spending.

Delaying or not paying value-added tax refunds diminishes the production efficiency of the VAT. It could force firms to substitute away from taxed to untaxed inputs such as labor. Exporters with a higher share of intermediate inputs generate large refund requests and would be most vulnerable to slow payment of VAT refunds. Using panel regressions at the industry-country level, Sharma (2020) studies the response of exporters to the adoption of the VAT, finding that a 10 percent higher

industry-level share of intermediate inputs is associated with an 8 percent decline in exports originating from that industry after the VAT's adoption.

In many value-added tax systems, the alternative to claiming a VAT refund is to “carry forward” the negative VAT liability. The idea here is that the negative VAT liability will be used as a tax credit against the firm's future VAT liabilities. Most countries actually *require* nonexporting firms to carry forward their negative VAT liability for several periods or until the amount becomes sufficiently high before they can claim the outstanding amount as refund. This policy is justified by the desire to prevent firms with relatively small tax credits from overwhelming the refund system. However, the result is that in some countries refunds are available (either *de facto* or *de jure*) only to exporters or large firms. Yet even when firms are allowed to request a refund rather than carry tax credits forward, they may choose the latter option if the refund process is believed to be slow, to require costly compliance actions, or to be associated with a higher audit probability than carry forward.

In our administrative data, carry-forward tax credits are captured in a consistent manner, allowing us to examine variation within and across countries. In Figure 4, panel B, we show that the share of firms carrying forward credits for at least twelve consecutive months is higher in lower-income countries, indicating a more pronounced departure from the textbook value-added tax system in which substantial carry forwards must not arise. In Hungary and Costa Rica, where the VAT refund system is relatively quick (ten weeks in Hungary, according to World Bank data), the share of firms that are carrying forward credits for more than twelve consecutive months is almost zero. The share is much more substantial in Pakistan, Senegal, and Ethiopia, where between 20 and 40 percent of all firms have been carrying forward credits for more than twelve months. The amounts carried forward are quantitatively important. On average, across the countries in the sample, the value of credits carried forward is equivalent to 42 percent of total annual sales (Figure 4, panel C). In all countries except Ethiopia, smaller firms carry forward a higher share of credits relative to sales, including even high-income countries such as France and Hungary.

The implications of the prevalence of carry forwards for the production and revenue efficiency of the value-added tax depend on whether firms have accumulated their tax credit legitimately or fraudulently. If negative tax liabilities are legitimate, our findings imply that real-world VAT systems generate cash constraints for a much wider group of firms than previously thought. The discussion surrounding VAT refunds tends to focus on exporters (Slemrod and Velayudhan 2022; Sharma 2020), but only 1–2 percent of firms export a substantial share of their sales, compared to the 20–40 percent of firms that carry forward credits for more than twelve months. On the other hand, if these firms' negative tax liabilities are generated illegitimately, our findings suggest that tax authorities may need to allocate some audit resources away from refund requests to the large number of firms that carry forward credits.

Most countries in our sample have provisions for large exporters to be “fast-tracked” for value-added tax refunds. This is presumably because refund claims by large exporters are more likely to be legitimate, as they are under tighter monitoring from the tax administration (Basri et al. 2021; Bachas, Fattal Jaef, and Jensen

2019). Also, misreporting domestic sales as exports is difficult, due to the third-party reporting of exports by customs. Despite these fast-track provisions, exporters (defined as having annual reported exports greater than 30 percent of annual sales) are typically more, not less, likely to be carrying forward credits for more than twelve months, compared to other firms. This pattern suggests either that the refund fast-track for exporters is not effective or that exporters do not use this option, choosing instead to carry forward the accumulated credits.

Overall, the fact that firms face delays in obtaining refunds constitutes an important departure from the textbook value-added tax system—a departure that is again more pronounced in lower-income countries. Managing the VAT refund system is often one of the biggest challenges for tax authorities in lower-income countries.

### **Should the Value-Added Tax Be Replaced by a Retail Sales Tax or a Turnover Tax?**

Given these numerous departures of real-world value-added tax systems from the textbook model and the real-world complexities involved in administering the VAT, several countries have recently considered replacing it. For example, Zambia briefly considered in 2019 replacing its VAT with a turnover tax largely due to problems with managing refunds (as reported in Asquith 2019), Malaysia actually replaced the VAT with a turnover tax in 2018 (Avalara n.d.), and Ghana introduced a sales tax on top of the VAT in 2018, partly to reduce the likelihood that firms will require a refund. The closest alternatives to a broad-based VAT are the retail sales tax and the turnover tax, which applies to the sales of firms at all levels of production. However, we will argue that these alternative taxes seem unlikely to perform better than the VAT in lower-income countries.

A textbook retail sales tax should be equivalent in both revenue and incidence to a textbook value-added tax. To see this, consider our example of a farmer-mill-bakery supply chain again and assume that the whole chain is in the VAT net with no exemptions or tax evasion. In a retail sales tax, the entire tax would be remitted by the bakery alone on its sales to final consumers. In contrast, in a VAT the tax would be remitted by each business in the supply chain in proportion to their value added, and the total amount of tax would be exactly the same that the government would receive under a retail sales tax. In fact, absent any enforcement concerns, a retail sales tax would arguably be preferable to a VAT, as it raises the same amount of revenue at lower compliance costs.

We use transaction-level data from Pakistan to simulate the counterfactual retail sales tax revenue the country would collect if it were to apply a retail sales tax at the same rate as the value-added tax.<sup>10</sup> To estimate the retail sales tax base, we

<sup>10</sup>Although we have access to transaction-level data for Rwanda and Uganda, we are not able to identify firms in the retail sector in a precise way in these data. We hence conduct the retail sales tax simulation using the Pakistan data only, as they are most suitable for the purpose. For details of the calculations

sum all sales reported by retail firms where the other party to the transaction is not a VAT-registered firm (that is, either final consumers or unregistered firms). We then apply the standard value-added tax rate to this base to get the counterfactual retail sales tax revenue. However, this calculation finds that the retail sales tax would raise at most one-third of VAT revenues in Pakistan. Why? The retail sector is typically less tax-compliant than other sectors in the economy, because sales to final consumers are reported to the government by one side only, in contrast to other transactions that are reported separately by both the seller and the buyer. Conversely, upstream stages in the supply chain (like manufacturers and importers) are relatively more tax-compliant, as upstream firms are larger and more tightly monitored by the tax administration. In effect, collecting VAT from the upstream firms is a form of tax withholding that does not exist in the retail sales tax (Waseem 2022).

Our basic calculations are probably too optimistic for the retail sales tax. Once a value-added tax was replaced with a retail sales tax, the paper trails flowing from upstream sectors to the retail sector would no longer exist, likely worsening the tax compliance in the retail sector even further. Moreover, although a retail sales tax could in theory tax services, it usually does not. This matters, because as economies develop, they undergo a structural transformation, shifting from the production of physical goods to services. For this reason, replacing a VAT with a retail sales tax in a lower-income country could be a step backwards, leading to erosion of the tax base as the country develops. Similarly, owing largely to difficulties in distinguishing between business-to-business and business-to-consumer transactions, a retail sales tax often ends up taxing business inputs despite not being designed to do so. For example, consider a baking soda sold to a consumer (business-to-consumer transaction) and sold to a bakery (business-to-business transaction). In principle, the retail sales tax should apply only to the first transaction. But for all practical purposes, it is not feasible to distinguish between the two transactions with the consequence that the retail sales tax also applies to an intermediate business input. This tax cascading, in which a tax is applied repeatedly to the gross value of sales along a supply chain, results in effective tax rates that may vary in haphazard ways, adding another layer of inefficiency into the tax system.<sup>11</sup> In short, while a retail sales tax is nondistortionary in basic public finance theory, it is not necessarily a welfare-maximizing policy in lower-income countries.

What about replacing the value-added with a turnover tax—a tax applied to all sales, both intermediate and final? A turnover tax is broad based and potentially straightforward to collect. The main concern is that because such a tax is applied to gross sales through the supply chain, not to value added, the tax rate would be applied multiple times, leading to tax cascading and creating a wedge between the prices of taxed and untaxed inputs. However, because there is no adjustment of tax anywhere in the supply chain, a turnover tax could theoretically be applied at a

---

described in this section comparing a retail sales tax and a turnover tax to a value-added tax, see the online Appendix.

<sup>11</sup> For a deeper discussion of the issues created by retail sales tax, see Fox (2012).



lower rate while raising the same amount as the current VAT. Our own calculations based on data from our eleven-country sample suggest that the revenue-neutral turnover tax rate varies across countries but is more than one-half of the standard VAT rate for most countries in our sample. The typical revenue-neutral turnover tax rate in our sample is about 5–8 percent. Note that a turnover tax would be imposed on imports, and because imports are relatively easy for tax officials to track through customs data, it would tend to put imports at a disadvantage. Any good with a longer supply chain across firms would also be at a disadvantage, thus distorting production. There would be a tax incentive for firms to integrate vertically, to avoid the cascade of the turnover tax. Although we are not aware of any empirical study that estimates welfare losses from tax cascading, simulations in Keen (2014) show that such losses could be sizable. Indeed, the welfare losses from cascading and vertical integration are likely to be first order and can easily dwarf any gains from the lower compliance costs of a turnover tax. In fact, the distortions generated by a broad-based turnover tax might be so large that it might not even be possible to recover the VAT revenue with a turnover tax.

Many countries do use turnover taxes for small taxpayers, often allowing small firms to opt for the simplicity of a turnover tax, generally in lieu of either a value-added tax or a corporate income tax. For the set of countries in the available IMF data (Wei and Wen 2019), the average turnover tax rate is 3.6 percent. The distortionary effects of a turnover tax for small taxpayers only are likely to be substantially less than those of a broad-based turnover tax because there is limited opportunity for the tax to cascade, especially if small taxpayers are more concentrated towards the retail end of the production chain.

In sum, our calculations with respect to a retail sales tax and a turnover tax suggest that, despite the practical challenges of a value-added tax, especially in lower-income countries, it still dominates the alternatives in terms of both production and revenue efficiency.

## **Discussion**

In an ideal world, a textbook value-added tax is an efficient way to raise revenue. In the real world, actual VAT systems fall short of this ideal, especially in lower-income economies. Revenue is heavily concentrated among a few large firms. Exemptions and reduced rates mean that effective tax rates vary substantially across products and firms, even within the same industry. Nonclaiming of input VAT by small firms suggests that the VAT chain is highly fragmented. Slow and ineffective VAT refunds constrain the liquidity of many firms, including exporters, effectively widening the wedge between the prices of taxed and untaxed goods. Our within-country analyses suggest that small firms appear as if they report VAT in a way that disadvantages them, for example by not claiming input VAT on legitimate purchases or not claiming exemptions and reduced rates when larger firms in

the same industry do claim them. At the same time, other studies have shown that smaller firms are much more likely to evade (Best, Shah, and Waseem 2022).

The outcomes of real-world value-added taxes that we document in this paper are due to a combination of policy choices, the administrative implementation of those policies, and the structural features of the economies in which VATs are implemented. An underlying theme is that real-world VAT systems display considerable heterogeneity, even in countries with similar levels of GDP per capita. For example, standard VAT rates are broadly similar across countries, bar a few outliers, and uncorrelated with countries' income levels, both in our sample and in a broader cross-section of countries. Exemptions and reduced rates lead to a slightly larger VAT revenue loss in lower-income countries compared to higher-income countries, but this difference is neither economically nor statistically significant. Instead, differences in VAT exemptions and reduced rates between countries with similar income per capita are much more substantial than differences across income levels. Moreover, real-world VAT systems seem to also change frequently, as, for example, new rate categories, exemptions, or administrative procedures are introduced. This creates uncertainty for firms, all the more so as policymakers use the VAT system not only as a tax instrument but also as a tool for redistribution and fiscal stimulus.

The variations of real-world value-added tax systems provide considerable scope for research on the determinants and outcomes of VATs, as this paper has suggested. One topic for future work is to examine the extent to which small firms leave money on the table while simultaneously engaging in tax evasion, or whether there are multiple types of small firms whereby some of them are naïve—not claiming input VAT, exemptions, or reduced rates—and others are sophisticated evaders (as suggested in Almunia et al. 2022). Such an analysis could provide a better understanding of the equity impact of the VAT and its features.

Although the value-added tax in practice is far less production efficient and revenue efficient than the textbook model, especially in lower-income countries, replacing it with alternative indirect taxes would create serious problems. The main alternatives, namely a retail sales tax or a turnover tax, would either not raise nearly as much revenue as the VAT or would do so at a high cost to production efficiency.

Thus, improving the functioning of value-added taxes in the real world offers a potentially more productive agenda. VAT withholding or “reverse charging,” for example, may limit tax evasion by small firms by mandating larger firms, government, or payment providers to remit tax on behalf of small firms. Withholding has not been extensively discussed in this paper due to the inconsistent reporting of withholding across countries, but an emerging literature has documented its positive impact on compliance (Brockmeyer and Hernandez 2022; Garriga and Tortarolo 2022). Effective fast-track mechanisms for refunds are another system tweak that could substantially improve the functioning of real-world VATs. Finally, thinking about how to optimize the use of third-party and digitally-reported data for enhancing VAT systems in a way that takes into account both spillovers and general

equilibrium effects in the production network are important tasks for both policy-makers and researchers.

■ We thank audiences at CIFREL, IFS, IIPF, Mannheim, the Oxford CBT, Tampere, UC3M, SAEe, the World Bank and Zurich, and Stuart Adam, David Agrawal, Pierre Bachas, Youssef Benzarti, Mick Keen, Helen Miller, Joana Naritomi, David Phillips, Oyebola Okunogbe, Joel Slemrod, and Dario Tortarolo for useful comments. We are exceedingly grateful to the staff of revenue authorities in our partner countries who facilitated access to the data and supported the analysis with excellent feedback and technical inputs. We also thank Bálint Ván, Stefano D'Angelo, Pablo García-Guzmán, Oliver Hanney, Adrienne Lees, Fabrizio Santoro, Thiago Scot, and Péter Tóth for generous support in implementing the analysis in select countries. We gratefully acknowledge funding from Fundación Ramón Areces, the International Centre for Tax and Development, from UK aid from the UK government through the Centre for Tax Analysis in Developing Countries (TaxDev) and from the UKRI via Brockmeyer's Future Leaders Fellowship (grant reference MR/V025058/1). The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of the World Bank, its Board of Executive Directors, or the governments that they represent.

## References

- AEAT. 2019. "Official Value-Added Tax Statistics for Fiscal Year 2019." Agencia Estatal de Administración Tributaria. [https://sede.agencia tributaria.gob.es/AEAT/Contenidos\\_Comunes/La\\_Agencia\\_Tributaria/Estadisticas/Publicaciones/sites/ivapartidas/2019/jrubikf2df48a69dfb61b75d414023bcf39a8d986e19e52.html](https://sede.agencia tributaria.gob.es/AEAT/Contenidos_Comunes/La_Agencia_Tributaria/Estadisticas/Publicaciones/sites/ivapartidas/2019/jrubikf2df48a69dfb61b75d414023bcf39a8d986e19e52.html) (accessed October 30, 2023).
- Agrawal, David R., and David E. Wildasin. 2020. "Technology and Tax Systems." *Journal of Public Economics* 185: 104082.
- Agrawal, David R., and Laura Zimmermann. 2019. "Production and Evasion Responses with Limited State Capacity." International Growth Centre Working Paper S-89411-INC-1.
- Agrawal, David R., and William F. Fox. 2021. "Taxing Goods and Services in a Digital Era." *National Tax Journal* 74 (1): 257–301.
- Almunia, Miguel, Jonas Hjort, Justine Knebelmann, and Lin Tian. 2022. "Strategic or Confused Firms? Evidence from 'Missing' Transactions in Uganda." *Review of Economics and Statistics*. [https://doi.org/10.1162/rest\\_a\\_01180](https://doi.org/10.1162/rest_a_01180).
- Asquith, Richard. 2019. "Zambia Stops Replacement of VAT with Sales Tax." Avalara, September 28. <https://www.avalara.com/vatlive/en/vat-news/zambia-stops-replacement-of-vat-with-sales-tax.html>.
- Atkinson, Anthony Barnes, and Joseph E. Stiglitz. 1976. "The Design of Tax Structure: Direct versus Indirect Taxation." *Journal of Public Economics* 6 (1–2): 55–75.
- Avalara. n.d. "Malaysian Sales and Service Tax." <https://www.avalara.com/vatlive/en/country-guides/asia/malaysia.html> (accessed on October 30, 2023).
- Bachas, Pierre, Anne Brockmeyer, Roel Dom, and Camille Semelet. 2023. "Effective Tax Rates and Firm Size." CEPR Discussion Paper 17985.

- Bachas, Pierre, Anne Brockmeyer, Alipio Ferreira, and Bassirou Sarr.** 2022. "How to Target Enforcement at Scale? Evidence from Tax Audits in Senegal." Economic Development and Institutions Working Paper.
- Bachas, Pierre, Anne Brockmeyer, and Camille Semelet.** 2020. "The Impact of COVID-19 on Formal Firms: Micro Tax Data Simulations across Countries." World Bank Policy Research Working Paper 9437.
- Bachas, Pierre, Lucie Gadenne, and Anders Jensen.** 2020. "Informality, Consumption Taxes and Redistribution." NBER Working Paper 27429.
- Bachas, Pierre, Roberto N. Fattal Jaef, and Anders Jensen.** 2019. "Size-Dependent Tax Enforcement and Compliance: Global Evidence and Aggregate Implications." *Journal of Development Economics* 140: 203–22.
- Basri, M. Chatib, Mayara Felix, Rema Hanna, and Benjamin A. Olken.** 2021. "Tax Administration versus Tax Rates: Evidence from Corporate Taxation in Indonesia." *American Economic Review* 111 (12): 3827–71.
- Baunsgaard, Thomas, and Michael Keen.** 2010. "Tax Revenue and (or?) Trade Liberalization." *Journal of Public Economics* 94 (9–10): 563–77.
- Benzarti, Youssef, Dorian Carloni, Jarkko Harju, and Tuomas Kosonen.** 2020. "What Goes Up May Not Come Down: Asymmetric Incidence of Value-Added Taxes." *Journal of Political Economy* 128 (12): 4438–74.
- Besley, Timothy, and Torsten Persson.** 2014. "Why Do Developing Countries Tax So Little?" *Journal of Economic Perspectives* 28 (4): 99–120.
- Best, Michael Carlos, Anne Brockmeyer, Henrik Jacobsen Kleven, Johannes Spinnewijn, and Mazhar Waseem.** 2015. "Production versus Revenue Efficiency with Limited Tax Capacity: Theory and Evidence from Pakistan." *Journal of Political Economy* 123 (6): 1311–55.
- Best, Michael, Jawad Shah, and Mazhar Waseem.** 2022. "Detection without Deterrence: Tax Audits with Limited Fiscal Capacity." Unpublished.
- Bird, Richard, and Pierre-Pascal Gendron.** 2007. *The VAT in Developing and Transitional Countries*. Cambridge, UK: Cambridge University Press.
- Brautigam, Deborah, Odd-Helge Fjeldstad, and Mick Moore, ed.** 2008. *Taxation and State-Building in Developing Countries: Capacity and Consent*. Cambridge, UK: Cambridge University Press.
- Brockmeyer, Anne, Giulia Mascagni, Vedanth Nair, Mazhar Waseem, and Miguel Almunia.** 2023. "Replication data for: Does the Value Added Tax Add Value? Lessons Using Administrative Data from a Diverse Set of Countries." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E195161V1>.
- Brockmeyer, Anne, and Magaly Sáenz Somarriba.** 2022. "Electronic Payment Technology and Tax Compliance: Evidence from Uruguay's Financial Inclusion Reform." CEPR Discussion Paper 17097.
- Brockmeyer, Anne, and Marco Hernandez.** 2022. "Taxation, Information, and Withholding: Evidence from Costa Rica." CEPR Discussion Paper 17716.
- Brockmeyer, Anne, Spencer Smith, Marco Hernandez, and Stewart Kettle.** 2019. "Casting a Wider Tax Net: Experimental Evidence from Costa Rica." *American Economic Journal: Economic Policy* 11 (3): 55–87.
- Brown, Charles, and James Medoff.** 1989. "The Employer Size-Wage Effect." *Journal of Political Economy* 97 (5): 1027–59.
- Caragher, Jacinta.** 2023. "How many countries have VAT or GST? 175." <https://www.vatcalc.com/global/how-many-countries-have-vat-or-gst-174/> (accessed on November 30, 2023).
- Carrillo, Paul, Dave Donaldson, Dina Pomeranz, and Monica Singhal.** Forthcoming. "Ghosting the Tax Authority: Fake Firms and Tax Fraud in Ecuador." *American Economic Review: Insights*.
- Carrillo, Paul, Dina Pomeranz, and Monica Singhal.** 2017. "Dodging the Taxman: Firm Misreporting and Limits to Tax Enforcement." *American Economic Journal: Applied Economics* 9 (2): 144–64.
- Coolidge, Jacqueline.** 2012. "Findings of Tax Compliance Cost Surveys in Developing Countries." *eJournal of Tax Research* 10 (2): 250–87.
- Crandall, William, Elizabeth Gavin, and Andrew Masters.** 2021. "ISORA 2018: Understanding Revenue Administration." International Monetary Fund Departmental Paper DP/2021/025.
- Crawford, Ian, Michael Keen, and Stephen Smith.** 2010. "Value Added Tax and Excises." In *Dimensions of Tax Design: The Mirrlees Review*, edited by Stuart Adam, Timothy Besley, Richard Blundell, Stephen Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba, 275–362. Oxford, UK: Oxford University Press.

- de Paula, Áureo, and Jose A. Scheinkman.** 2010. "Value-Added Taxes, Chain Effects, and Informality." *American Economic Journal: Macroeconomics* 2 (4): 195–221.
- Diamond, Peter A., and James A. Mirrlees.** 1971. "Optimal Taxation and Public Production I: Production Efficiency." *American Economic Review* 61 (1): 8–27.
- Direction Générale des Impôts et des Domaines.** 2015. "VAT Declaration Data for Senegal, 2009–2015." Unpublished confidential dataset.
- Ebrill, Liam, Michael Keen, Jean-Paul Bodin, and Victoria Summers.** 2001. *The Modern VAT*. Washington, DC: International Monetary Fund.
- eSwatini Revenue Service.** 2018. "VAT Declaration Data for eSwatini, 2012–2017." Unpublished confidential dataset.
- Federal Board of Revenue.** 2018. "VAT Declaration Data for Pakistan, 2012–2018." Unpublished confidential dataset.
- Fisman, Raymond, and Shang-Jin Wei.** 2004. "Tax Rates and Tax Evasion: Evidence from 'Missing Imports' in China." *Journal of Political Economy* 112 (2): 471–96.
- Fox, William F.** 2012. "Retail Sales and Use Taxation." In *The Oxford Handbook of State and Local Government Finance*, edited by Robert D. Ebel and John E. Petersen, 406–28. Oxford, UK: Oxford University Press.
- Gadenne, Lucie, Tushar K. Nandi, and Roland Rathelot.** 2022. "Taxation and Supplier Networks: Evidence from India." IFS Working Paper W19/21.
- Garriga, Pablo, and Dario Tortarolo.** 2022. "Firms as Tax Collectors." IFS Working Paper 22/44.
- Gerard, François, Joana Naritomi, Arthur Seibold, and Bruno Zulian.** 2022. "Two-Tier Tax Systems and Firms: Evidence from Brazil." Unpublished.
- Highfield, Richard, Chris Evans, Binh Tran-Nam, and Michael Walpole.** 2019. "Diagnosing the VAT Compliance Burden: A Cross-Country Comparison." <http://doi.org/10.26190/5de04d9259dc1>.
- Her Majesty's Revenue and Customs (HMRC).** 2018–2019. "Official Statistics: Value-Added Tax Statistics for Fiscal Year 2018/19." Gov.UK. <https://www.gov.uk/government/statistics/value-added-tax-vat-annual-statistics> (accessed on October 30, 2023).
- Hoy, Christopher.** 2022. "How Does the Progressivity of Taxes and Government Transfers Impact People's Willingness to Pay Tax?" World Bank Policy Research Working Paper 10167.
- Iddrisu, Abdul Malik, Harshil Parekh, and David Phillips.** 2023. "How a Tax Exemption Can Increase Revenues." TaxDev, March 17. <https://www.taxdev.org/news-events/how-tax-exemption-can-increase-revenues>.
- Keen, Michael.** 2007. "VAT Attacks!" *International Tax and Public Finance* 14 (4): 365–81.
- Keen, Michael.** 2008. "VAT, Tariffs, and Withholding: Border Taxes and Informality in Developing Countries." *Journal of Public Economics* 92 (10–11): 1892–1906.
- Keen, Michael.** 2014. "Targeting, Cascading and Indirect Tax Design." *Indian Growth and Development Review* 7 (2): 181–201.
- Keen, Michael, and Jack Mintz.** 2004. "The Optimal Threshold for a Value-Added Tax." *Journal of Public Economics* 88 (3–4): 559–76.
- Liu, Li, Ben Lockwood, Miguel Almunia, and Eddy H. F. Tam.** 2021. "VAT Notches, Voluntary Registration, and Bunching: Theory and U.K. Evidence." *Review of Economics and Statistics* 103: 151–64.
- Mascagni, Giulia, Roel Dom, Fabrizio Santoro, and Denis Mukama.** 2023. "The VAT in Practice: Equity, Enforcement, and Complexity." *International Tax and Public Finance* 30 (2): 525–63.
- Mascagni, Giulia, Andualem T. Mengistu, and Firew B. Woldeyes.** 2021. "Can ICTs Increase Tax Compliance? Evidence on Taxpayer Responses to Technological Innovation in Ethiopia." *Journal of Economic Behavior and Organization* 189: 172–93.
- Mascagni, Giulia, Fabrizio Santoro, Denis Mukama, John Karangwa, and Naphthal Hakizimana.** 2022. "Active Ghosts: Nil-filing in Rwanda." *World Development* 152: 105806.
- Ministerio de Hacienda.** 2015. "VAT Declaration Data for Costa Rica, 2007–2014." Unpublished confidential dataset.
- Ministère des Finances (DGFIP).** 2021. «TVA\_annuel: Données détaillées déclaratives annuelles de TVA.» Confidential dataset.
- Ministero de Finanzas Publicas.** 2022. "Ingresos Tributarios de la Administración Central 2002–22." <https://www.minfin.gob.gt/images/daf/documentos/doc108.pdf> (accessed on November 11, 2023).
- Ministry of Revenue.** 2018. "VAT Declaration Data for Ethiopia, 2011–2017." Unpublished confidential dataset.

- Mirrlees, James, Stuart Adam, Tim Besley, Richard Blundell, Stephen Bond, Robert Chote, Malcolm Gammie, Paul Johnson, Gareth Myles, and James Poterba. 2011. *Tax by Design: The Mirrlees Review*. Oxford, UK: Oxford University Press.
- Naritomi, Joana. 2019. "Consumers as Tax Auditors." *American Economic Review* 109 (9): 3031–72.
- Pénzügyminisztérium. 2021. "VAT Declaration Data for Hungary, 2017–2019." Unpublished confidential dataset.
- Pessoa, Mario, Andrew Kazora Okello, Artur Swistak, Muyangwa Muyangwa, Virginia Alonso-Albarran, and Vincent Koukpaizan. 2021. "How to Manage Value-Added Tax Refunds." International Monetary Fund How to Notes 21/04.
- Pomeranz, Dina. 2015. "No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax." *American Economic Review* 105 (8): 2539–69.
- Poschke, Markus. 2018. "The Firm Size Distribution across Countries and Skill-Biased Change in Entrepreneurial Technology." *American Economic Journal: Macroeconomics* 10 (3): 1–41.
- Ramsey, F. P. 1927. "A Contribution to the Theory of Taxation." *Economic Journal* 37 (145): 47–61.
- Redonda, Agustín, Christian von Haldenwang, and Flurim Aliu. 2022. "Global Tax Expenditures Database (GTED)." <https://doi.org/10.5281/ZENODO.6334212> (accessed on October 30, 2023).
- Rwanda Revenue Authority. 2020. "VAT Declaration Data for Rwanda, 2016–2020." Unpublished confidential dataset.
- Servicio de Administración de Rentas. 2020. "VAT Declaration Data for Honduras, 2018–2020." Unpublished confidential dataset.
- Sharma, Rishi R. 2020. "Does the VAT Tax Exports?" *Economic Inquiry* 58 (1): 225–40.
- Slemrod, Joel, Brett Collins, Jeffrey L. Hoopes, Daniel Reck, and Michael Sebastiani. 2017. "Does Credit-Card Information Reporting Improve Small-Business Tax Compliance?" *Journal of Public Economics* 149: 1–19.
- Slemrod, Joel, and Tejaswi Velayudhan. 2022. "The VAT at 100: A Retrospective Survey and Agenda for Future Research." *Public Finance Review* 50 (1): 4–32.
- Superintendencia de Administración Tributaria. 2020. "VAT Declaration Data for Guatemala, 2006–2015." Unpublished confidential dataset.
- Tait, Alan A. 1988. *Value Added Tax: International Practice and Problems*. Washington, DC: International Monetary Fund.
- Tourek, Gabriel. 2022. "Targeting in Tax Behavior: Evidence from Rwandan Firms." *Journal of Development Economics* 158: 102911.
- Uganda Revenue Authority. 2018. "VAT Declaration Data for Uganda, 2011–2018." Unpublished confidential dataset.
- Ulysea, Gabriel. 2018. "Firms, Informality, and Development: Theory and Evidence from Brazil." *American Economic Review* 108 (8): 2015–47.
- UNU-WIDER. 2023. "UNU-WIDER Government Revenue Dataset 2023." <https://doi.org/10.35188/UNU-WIDER/GRD-2023> (accessed on November 30, 2023).
- Warwick, Ross, Tom Harris, David Phillips, Maya Goldman, Jon Jellema, Gabriela Inchauste, and Karolina Goraus-Tańska. 2022. "The Redistributive Power of Cash Transfers vs VAT Exemptions: A Multi-country Study." *World Development* 151: 105742.
- Waseem, Mazhar. 2022. "The Role of Withholding in the Self-Enforcement of a Value-Added Tax: Evidence from Pakistan." *Review of Economics and Statistics* 104 (2): 336–54.
- Waseem, Mazhar. 2023. "Overclaimed Refunds, Undeclared Sales, and Invoice Mills: Nature and Extent of Noncompliance in a Value-Added Tax." *Journal of Public Economics* 218: 104783.
- Wei, Feng, and Jean-François Wen. 2019. "The Optimal Turnover Threshold and Tax Rate for SMEs." International Monetary Fund Working Paper WP/19/98.
- World Bank. 2020. *Doing Business 2020: Comparing Business Regulation in 190 Economies*. Washington, DC: World Bank.
- World Bank. 2023. "World Development Indicators." Dataset. Available at: <https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators> (accessed on November 8, 2023).
- Yesegat, Wollela Abehodie, Denis Vorontsov, Jacqueline Coolidge, and Laurent Olivier Corthay. 2016. *Tax Compliance Cost Burden and Tax Perceptions Survey in Ethiopia*. Washington, DC: World Bank.

# The Failure of Silicon Valley Bank and the Panic of 2023

Andrew Metrick

**O**n Thursday, March 9, 2023, depositors withdrew almost 25 percent of the total deposits at Silicon Valley Bank (SVB). On the morning of Friday, March 10, early evidence suggested that virtually all of the remaining deposits would be withdrawn. Regulators were forced to close the bank in the middle of the day, unable to even “get to the weekend” as had been done for every other bank failure since the Great Depression. Silicon Valley Bank, with \$209 billion of assets, the sixteenth-largest bank in the United States, became the second-largest bank failure in nominal terms in US history, ranking in size only behind the failure of Washington Mutual Bank during the Great Recession of 2007–2009.

Since the founding of the Federal Deposit Insurance Corporation (FDIC) in 1934, no US bank depositor has lost any insured funds. As a result, US bank runs have been rare, because depositors see no need to rush to the exits, even if a bank failure is imminent. However, US deposit insurance has a limit on the amount insured of \$250,000 per account. Silicon Valley Bank was highly unusual in that about 94 percent of its deposits are large enough to be *uninsured*. This makes SVB a throwback, almost as though it was from the bank-panic frenzies of the late nineteenth century, before the modern safety net for banks was in place. It also makes the failure of SVB an illustrative model for the economics of banking—how things work can often become most clear when they break.

After the failure of Silicon Valley Bank, commentary focused on several possible causes: poor risk management by the bank; changes in bank regulation

■ *Andrew Metrick is the Janet L. Yellen Professor of Finance and Management, Yale School of Management, New Haven, Connecticut. His email address is [andrew.metrick@yale.edu](mailto:andrew.metrick@yale.edu).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.1.133>.

in 2019 that had recently relaxed the rules for banks like SVB; weak oversight by supervisors for the rules that remained; a customer base highly concentrated in one industry (technology, particularly startups) and region (Silicon Valley, of course); and the interaction of modern social media with the tech-savvy customer base of SVB in accelerating the run. From this long list, one might conclude that SVB was a special case, an outlier that reveals little about bank stress in general. This would not, however, be correct, because the underlying problem was systemic, and when SVB was shut down, the Panic of 2023 had begun.

On Sunday, March 12, the Federal Deposit Insurance Corporation took control of another failed institution, Signature Bank, which then became the third-largest bank failure in the history of the United States. This failure convinced authorities that the US banking system faced the possibility of widespread runs. In response, the FDIC and the Federal Reserve each invoked emergency provisions available under the law. The “systemic-risk exception” included in a 1991 law allows the FDIC to act immediately without taking time to seek out a least-cost method of resolving a bank failure: in this case, the FDIC announced that even uninsured depositors would be fully covered at both Silicon Valley Bank and Signature. In turn, the Federal Reserve invoked its power to take action during “unusual and exigent circumstances” that was included in a 1935 revision of the Federal Reserve Act and created a novel emergency-lending program with generous terms. Together, these policy actions succeeded in slowing the deposit outflows, which delayed but did not prevent the eventual failure of First Republic Bank in May. At \$213 billion in assets, First Republic was even larger than SVB, thus leapfrogging to become the “new” second-largest bank failure ever in the United States. Panic in the US banking sector also spread to Europe, providing the final nail in the coffin for the long-troubled Credit Suisse, which in April 2023 became the single largest bank failure in the world, ever.

Each of these failures had their own idiosyncratic reasons. But the US examples all shared two things: a very high percentage of uninsured deposits and significant “unrealized” losses on assets. That combination is a dangerous mix, and variations on it lie at the core of most financial panics. In most modern examples, such as the global financial crisis of 2008–2009, this dangerous mix is shrouded in layers of complexity, making it much harder to uncover its dynamics. But the special case of Silicon Valley Bank turns out to be a remarkably clean example, and thus can teach us a lot about the general case.

To elicit these lessons, we focus on underlying economics of the Silicon Valley Bank failure, leaving most of the discussion of government regulation until after the basic story is told. In the next section, we examine the balance sheet of SVB at the end of 2019 and thus before the pandemic, focusing on the largest categories of assets and liabilities. Like most banks, SVB’s profits are driven mostly by “net interest margin”—the difference in interest earned and paid for via its assets and liabilities—and this business model has come under pressure since the onset of the pandemic. In the following section, we illustrate this pressure using diagrams showing the components of net interest margin and show how those components changed during the pandemic. With this background of SVB’s balance sheet and



business model, we can then make sense of what happened at SVB and the ensuing Panic of 2023. In the penultimate section of the paper, we turn to the role of government bank regulation and oversight in the lead up to the crisis. The final section concludes with a discussion of the policy challenge in preventing future panics of this type.

## Silicon Valley Bank before COVID-19

The top panel of Figure 1 shows the balance sheet for Silicon Valley Bank on December 31, 2019. At this time, the bank had a total of \$70 billion of assets. The single largest item on the balance sheet was \$63 billion of deposits, which comprised the vast majority of the bank's funding. SVB was unusual in that almost all of these deposits—about 94 percent—were not covered by federal deposit insurance.

In the United States, individual bank accounts are insured by the Federal Deposit Insurance Corporation up to a limit of \$250,000. Balances below \$250,000 are “insured,” while all balances above \$250,000 are “uninsured.” This limit is high enough that the vast majority of all *accounts* are fully insured, but on a *dollar* basis, about 42 percent of the total deposits in the United States were above this limit, and thus uninsured (FDIC 2023c, p. 10). The specific choice of \$250,000 as the legal threshold is an attempt to balance the benefits (financial stability) and costs (excess risk-taking by banks) of insurance. The Panic of 2023 has reopened the debate on this threshold—a topic we discuss in the conclusion of the paper.

FDIC insurance is backed by a standing fund financed by insurance premiums charged to banks, with a credit line from the US Treasury and the full faith and credit of the United States standing as additional lines of defense. For insured depositors, the typical bank failure is seamless. Uninsured deposits do not have this explicit protection. Nevertheless, in most bank failures the FDIC has been able to transfer the entire deposit base to another bank, allowing the uninsured depositors to be made whole. Indeed, the *total* losses by uninsured depositors in the last 30 years has been less than \$300 million (FDIC 2023c, p. 22). With this history, it would be reasonable for uninsured bank depositors to feel relatively safe, even if that safety has not been legally guaranteed. This implicit safety supports the significant level of uninsured deposits in the US banking system. Most of these uninsured deposits are concentrated in the largest banks. Even among this group, however, Silicon Valley Bank stood out for having the highest percentage of uninsured deposits among all banks with assets of \$50 billion or more.

On the asset side of the Silicon Valley Bank balance sheet in 2019, the largest line item was “loans” at \$33 billion, followed by “securities” at \$28 billion. Most of the loans were in the form of credit lines to SVB customers, instead of more traditional term loans for specific projects; the startup firms served by SVB often have lumpy inflows and outflows and rely on credit lines to smooth those lumps. Very few of the loans were residential or commercial mortgages. The securities holdings were almost exclusively government or agency-backed bonds. With securities holdings at

Figure 1

**SVB Balance Sheet, 2019 and 2022**

Dec. 31, 2019 (\$billions)			
Assets		Liabilities + equity	
Cash	6	Deposits	63
Securities	28	Other debt	2
Loans	33		
Other	3	Total liabilities	65
		Equity	5
Total	70	Total	70

Dec. 31, 2022 (\$billions)			
Assets		Liabilities + equity	
Cash	13	Deposits	175
Securities	117	Other debt	19
<i>Mark-to-market</i>	<i>99</i>		
Loans	74		
Other	5	Total liabilities	194
		Equity	15
		<i>Mark-to-market</i>	<i>-3</i>
Total	209	Total	209
<i>Mark-to-market</i>	<i>191</i>	<i>Mark-to-market</i>	<i>191</i>

Source: SVB (2019, 2022).

Note: Mark-to-market equity represents author's calculations of adjusted (mark-to-market) assets less book liabilities.

40 percent of all assets, SVB was above the national average of 25 percent, but was not an extreme outlier. Banks hold securities for two main reasons: (1) as a higher-yielding (compared to cash) way to store funds that they plan to later loan out; and (2) as a source of liquidity, because safe government securities can easily be sold for cash to meet deposit withdrawals. Plain “cash” was only at \$6 billion.

Bank assets face two main risks: “credit risk” and “interest-rate risk.” Credit risk is that a borrower will not make timely payments of interest or principal; this risk is what usually gets banks into trouble, but was not a problem for Silicon Valley Bank. Their loan portfolio was concentrated in the ecosystem of startup technology firms, the venture capital firms that fund them, and the employees of both. Although this portfolio was not well-diversified, SVB never faced serious concerns about the quality of its loans. Furthermore, SVB’s securities portfolio was almost exclusively

comprised of government or agency-backed bonds, and thus had effectively zero credit risk.

The main problem at Silicon Valley Bank turned out to be interest-rate risk: in general, when interest rates change, the value of most assets and liabilities change as well, and this risk must be managed. Even risk-free securities like US government bonds are subject to interest-rate risk. For example, imagine that all interest rates are exactly zero, and you own a government security that will pay off in one year for exactly \$1, without making any other payments along the way. In this zero-rate world, the market value of this government security today should also be \$1. Next, imagine that all interest rates increase to be 1 percent. Now, if you wanted to sell that government security, any buyer will want to earn a return of 1 percent. For a bond that will ultimately pay \$1 in one year, the price today would need to be (approximately) 99 cents. Even though the payment of \$1 in one year is known for certain, the price of the security must still fall today to make the return (99 cents today turning into \$1 in one year) competitive with newly issued securities that offer a 1 percent return.

Interest-rate risk tends to grow with the maturity of a security. Consider a bond that has a single payment of \$1 in five years, with no other payments before that. In a zero-interest-rate world, the price of this bond today would still be \$1. But now, if interest rates increase to 1 percent, an investor today would only be willing to pay about 95 cents for that bond, because the investor would require a return of 1 percent per year (95 cents today turning into \$1 in five years) over the full five-year holding period. For the five-year bond, a 1 percentage point increase in interest rates leads to approximately a 5 percent decrease in the price. For the one-year bond example considered above, the decrease was only 1 percent.

With more complex bonds, the arithmetic gets messier, but the underlying intuition stays the same: when interest rates go up, the value of long-term assets is hurt more than short-term assets. If a bank wants to protect itself against interest-rate risk, it can try to match the maturity structure of its assets and liabilities, so that any change in interest rates would have an offsetting effect on both sides of the balance sheet. If this offset is perfect, then we would say that the bank has “perfectly hedged” its interest-rate risk.

Perfect hedges are difficult to achieve in banking. For most banks, the vast majority of their liabilities are customer deposits, most of which can be withdrawn upon demand. The legal maturity of such deposits is effectively zero. It is not feasible for traditional banks to get their assets down to zero maturity, so banks use a variety of other techniques to manage their interest-rate risk. The failure of Silicon Valley Bank to hedge its interest-rate risk properly was the proximate cause of its failure, but the story is more complicated than simple negligence. We return to this topic in the next section of the paper.

On a bank’s balance sheet, the difference between assets and liabilities is mechanically equal to “equity,” which in Silicon Valley Bank in 2019 stood at \$5 billion, yielding an equity-to-assets ratio of about 7 percent. For the purposes of this paper, we will consider equity to be a synonym for “capital”—that is, the

cushion that protects depositors and other debtholders from variations in the value of the assets. To the extent that assets are completely riskless in all respects, any positive level of capital would be sufficient to provide safety for all depositors, insured and uninsured. But once any uncertainty arises about the value of assets, a rational depositor (or insurer) would require more capital to feel safe.

One useful way to conceptualize the necessary capital cushion for a bank is to think of it as an input in the production of money. To satisfy the value of bank deposits (like checking accounts) as a transactions medium, it must be possible to use these deposits in exchange *at par value* with “no questions asked” (Holmstrom 2015). If an agent has to worry that the money they are receiving is not actually worth the number written on the check, then it will lose its usefulness as money. Having a sufficient capital buffer is the most important step a bank can take to ensure this monetary use. Conversely, if equity levels fall too low, then it would be reasonable for counterparties to be concerned about the quality of the liabilities, and the bank’s checks and deposits would lose their no-questions-asked property.

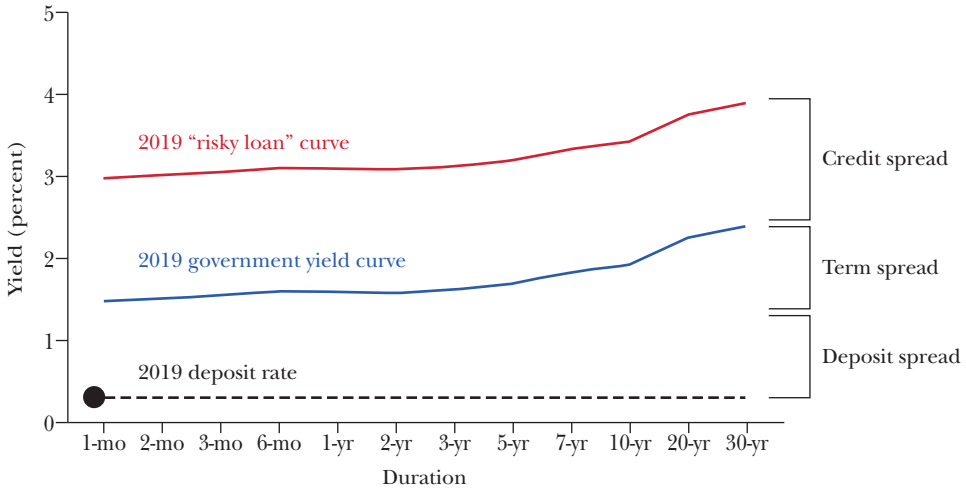
The link between the potential solvency of a bank (through its equity levels) and the liquidity of the bank (through the risk of bank runs) is driven by the connection of both to the no-questions-asked property of deposits. When bank money loses its no-questions-asked property, the rational response for (uninsured) depositors is to take their money out of the bank and put it someplace safer. Bank balance sheets contain many components that are costly to evaluate. No depositor earning a small (or zero) interest rate on their deposits should do a deep analysis of their bank to determine whether or not it was still safe—withdrawing deposits is much easier.

## **Silicon Valley Bank and the Business Model of Banking**

At its core, banking is a simple business. Almost all of a bank’s assets and liabilities are essentially just digital equivalents of pieces of paper: mortgages, commercial loans, checking accounts, cash reserves, and so on. The profitability of a bank is driven by earning higher interest on the assets than is paid on the liabilities, with this difference known as the “net interest margin.” Indeed, one can think of a bank of producing a consumer good (“deposit services”) that is itself just a liability, using the rest of its balance sheet as the factory floor for this production.<sup>1</sup>

<sup>1</sup>US banks earn about two-thirds of their revenue from net interest. The remaining noninterest revenue comes from fees and service charges on deposit accounts and credit cards, investment banking, asset management, and other activities. See Haubrich and Young (2019) for an analysis of these components over time. The Silicon Valley Bank business model focused mainly on net interest: “Net interest income accounts for the major portion of our earnings. It is comprised primarily of income generated from interest rate spread differences between the interest rates received on interest-earning assets, such as loans extended to clients and securities held in our fixed income securities portfolio, and the interest rates paid by us on interest-bearing liabilities, such as deposits and borrowings” (Silicon Valley Bank Financial Group 2023, p. 7).

Figure 2  
Silicon Valley Bank and the Business Model of Banking



Source: US Department of the Treasury (2023); SVB (2020a); author's calculations.

Notes: The 2019 deposit rate is a year-end estimate based on SVB presentations. The 2019 government yield curve represents the government's cost of borrowing at different maturities. The 2019 "risky loan" curve represents a 150-basis point premium on the government yield curve as an estimate of a premium on risky lending.

Figure 2 illustrates the three main sources of net interest margin for Silicon Valley Bank as of December 31, 2019. The y-axis shows interest rates (= "yield") measured as an annualized percentage; the x-axis shows time. Starting in the middle of the figure, the solid blue line shows the term structure of government debt: the market interest rates paid by the government to borrow at various maturities, ranging from one month to 30 years. (As is standard in these kinds of figures, the x-axis is not drawn to scale, instead showing many observations for short-term maturities before getting sparser at the long end.) The graph of these government interest rates, represented by the blue line, is called the "yield curve." The curve begins at approximately 1.5 percent for Treasury bills with one month remaining on their term, rising to about 2 percent for 10-year bonds and 2.5 percent for 30-year bonds. The "term spread" shown on the right side of the figure represents the difference between the yields on the shortest- and longest-horizon government debt. This term spread is positive, which is the usual condition. To the extent that SVB—like most banks—will be borrowing at the short end of the curve and lending at the long end of the curve, a bank will capture some component of this term spread as part of its net interest margin.

If all the bank did was to borrow short-term at the government rate and then lend long-term back to the government, the net interest margin would be exactly equal to the term spread. But banks can improve upon the seemingly risk-free

borrowing rates of the government. At the bottom left of the figure, the solid circle at 0.4 percent represents the average interest rate paid by Silicon Valley Bank on deposits at the end of 2019. Because these deposits can be demanded at any time, their *legal* term-to-maturity is close to zero, and thus the solid circle lies right next to the y-axis. The vertical difference between this solid circle and the leftmost point on the blue government yield curve is the “deposit spread” earned by SVB. Effectively, SVB’s customers are willing to accept an even lower return than they would get in government T-bills because of the transaction services and other perquisites provided by their demand deposits at the bank. Unlike Treasury bills, bank deposits can be used to facilitate transactions; many components of the legal ability to provide these services are unique to banks.

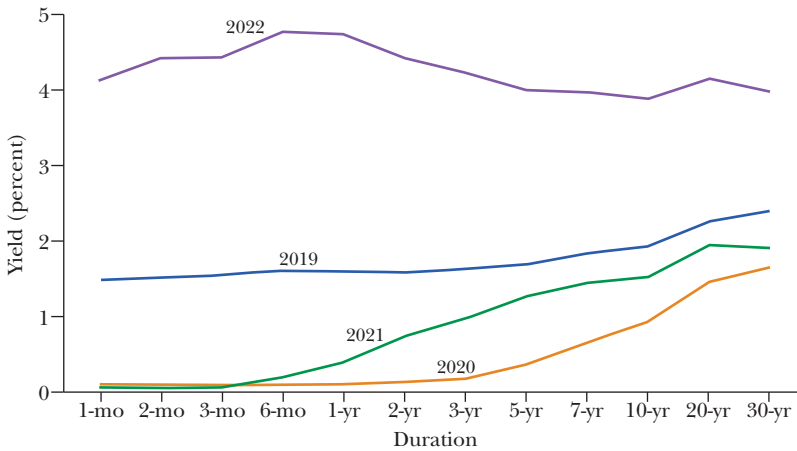
Figure 2 also includes a dotted line extended from the solid circle. This dotted line represents the expectation that deposits are “sticky” in the sense that depositors do not switch banks in the short run. In the limit, if we make an extreme (counterfactual) assumption that most depositors make a lifetime commitment to their bank at current dotted-line rates, then any further increase in the blue line would effectively be a windfall for the bank.

Of course, in normal times we do not expect banks to invest all of their deposits in government securities. As seen in the Silicon Valley Bank balance sheet in Figure 1, approximately half of the assets at SVB were in securities at the end of 2019, with the other half in more traditional loans to businesses and individuals. Unlike government securities, these traditional loans are risky and would be expected to earn a premium over the government yield curve. This “risky-loan” curve is shown in red, with the difference between the red and blue curves shown on the right as the “credit spread.” The risky-loan curve represents an estimate of expected yields on the loans, drawn just for expositional purposes: no public data is available to show the rates on the SVB loan book, so we just use a fixed premium here of 1.5 percent. Note that if a bank turns out to have chosen its loans poorly, the *realized* return on risky loans could easily lie beneath the government yield curve.

Taken together, the net interest margin in any given period would be the sum of these three components: the term spread, the deposit spread, and the credit spread. In ordinary times, when a bank is able to pay deposit rates below the government borrowing rate (positive deposit spread), when the government yield curve slopes up (positive term spread), and when risky loans are being paid back (positive realized credit spread), banks will have healthy net interest margins. For 2019, Silicon Valley Bank realized a net interest margin above 3 percent.

But the monetary policy response to COVID-19 led to major changes. Figure 3 traces the evolution of the government yield curve for the next three years. The blue line gives the starting point at the end of 2019, with the same curve as shown in the same color on Figure 2. Even though this curve was already low by historical standards, the response of the Federal Reserve to the pandemic drove rates even lower. The orange line shows the government yield curve at the end of 2020. Short-term rates were effectively zero, and Silicon Valley Bank—like all banks—was no longer earning a positive deposit spread.

Figure 3  
**Treasury Yield Curves, 2019–2022**



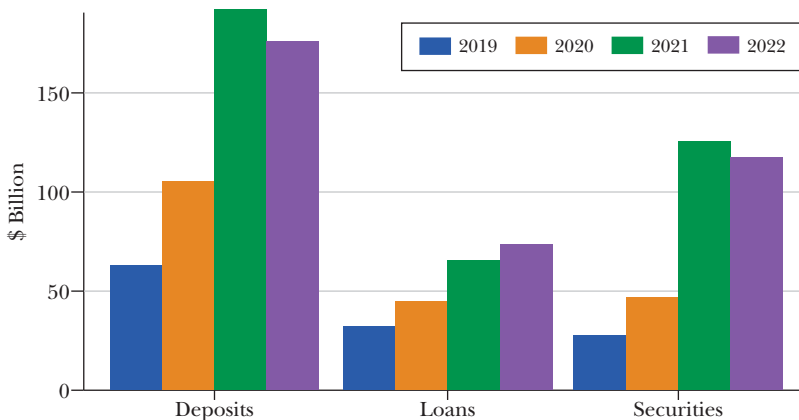
Source: US Department of the Treasury (2023).

Note: Lines represent the government’s cost of borrowing at different maturities, measured at year-end.

Had the yield curve stayed at these historical lows, Silicon Valley Bank would have been fine. Although the deposit spreads available at the end of 2019 had disappeared, the term spread still existed and the legacy loans from before the COVID era were continuing to perform, preserving credit spreads from that time. But the Fed began fighting the COVID-era inflation in earnest in early 2022, and by the end of the year the government yield curve (in purple) was strikingly different from its blue, orange, and green predecessors; both the level and the slope of the curve had changed. On December 31, 2022, the yield curve for government securities was relatively flat at about 4 percent for all maturities, eliminating the term-spread component.

Figure 4 illustrates the main changes to Silicon Valley Bank’s end-of-year balance sheet from 2019 to 2022, using the same color scheme as for interest rates in Figure 3. Taken together, these two figures show how the rapid growth at SVB coincided with changes in the interest-rate environment during the COVID era. In 2020, the powerful easing of rates was also combined with aggressive expansionary fiscal policy. The combined effect of these twin policies was felt particularly strongly in the tech sector of the Silicon Valley region. Younger firms were awash in cash, and a lot of that cash made its way to SVB, where deposits more than tripled from the end of 2019 to the end of 2021. Even in the best of times, it would have been difficult to find enough good lending opportunities to soak up these deposits. SVB did lend out some of these deposits, but the majority went into the purchase of securities. The same dynamics applied in 2021: the loan book grew, but not as fast as the securities portfolio.

Figure 4

**SVB: Loans, Deposits, and Securities, 2019–2022**

Source: SVB (2019, 2020b, 2021, 2022).

Note: This figure illustrates the main changes to Silicon Valley Bank's end-of-year balance sheet from 2019 to 2022, using the same color scheme as for interest rates in Figure 3.

The bank's growth in 2020 and 2021 took place in a low-interest-rate environment. When the Federal Reserve aggressively raised its target interest rate in 2022, the balance sheet at Silicon Valley Bank stopped growing. Towards the end of the year, tighter credit conditions and the need for operating funds led some of SVB's customers both to withdraw deposits and to tap their credit lines. To hold on to customers, SVB did begin to raise the interest rates paid on deposits, slightly, but even at the end of 2022, about half of all its deposits paid no interest.

For Silicon Valley Bank, the greatest damage of the interest-rate changes came from the effect on the balance sheet, as the increase in long-term interest rates reduced the value of all of the long-maturity assets. The bottom panel of Figure 1, showing the balance sheet on December 31, 2022, allows us to assess the damage. At this time, the book value (or original cost) of assets and liabilities was \$209 billion and \$194 billion respectively, leaving \$15 billion left over as an equity cushion. From this book-value accounting perspective, SVB was solvent and had an equity-to-assets ratio of approximately 7 percent, the same ratio as it had at the end of 2019.

But things look different once we "mark-to-market" the assets—that is, when we look not at the value of the assets at the time they were purchased, but at what they are worth at current market prices. For securities held by Silicon Valley Bank, most of which were purchased when interest rates were much lower in 2020 and 2021, the rise in long-term rates caused a reduction in mark-to-market value of about \$17 billion, with this newly mark-to-market value shown in red as \$103 billion. Note that these mark-to-market values are not based on realized or expected credit losses; rather, all of the reduction in value comes only from the fact that the value of a bond that pays a certain fixed interest rate will fall when market interest rates rise.



There is no mystery about this mechanical relationship, and it was well understood by management, bank supervisors, and market analysts. Indeed, most of these losses had already been incurred by the end of the third quarter of 2022, and the information about these losses was available in public filings by the bank well before the end of 2022. By this measure, Silicon Valley Bank was already insolvent by December 31, 2022, as shown in the bottom panel of Figure 1, with negative \$3 billion in equity. From these facts, one might wonder what regulators, investors, and uninsured depositors could possibly have been thinking to allow SVB to continue operating for so long after this apparent “insolvency” was clear.

The answer here is that Silicon Valley Bank relied on the stability of their customer relationships—the “stickiness” of deposits—as their main method to hedge interest-rate risk. At first glance, this belief may seem counterintuitive, and perhaps even a little magical. Most deposits are in-demand accounts and can be withdrawn at any time. In principle, if interest rates rise, one would expect that either deposit rates would rise or customers would leave and go to another bank. But reality is different. Loyalty, transactions costs, and simple inertia keep many consumers at their banks even when interest rates change—indeed, this stickiness is the reason for the dotted line in Figure 2.

Silicon Valley Bank is not unusual in relying on the stability of its deposits. In an influential paper, Dreschler, Savov, and Schnabl (2021) demonstrate that profits from the deposit spread have been a remarkably good hedge for interest-rate risk for US banks. Their analysis shows that deposit rates are quite inelastic to market interest rates, so that an increase in market rates leads to an increase in deposit *spreads*. Banks build their business plans around this relationship, using marketing, branch networks, and personal service to maximize the stability of their deposit base. We can think of the net present value of this deposit spread like an additional asset for banks, the “franchise value of deposits,” but one that is never included in any formal balance sheets. When interest rates rise, the deposit spread increases and this franchise value goes up, but not even “mark-to-market” accounting will capture this change.

The rise in interest rates in 2022 caused significant losses for the entire banking system (Jiang et al. 2023; Flannery and Sorescu 2023). Hundreds of banks were likely close to insolvency if we were to accurately mark-to-market *only the securities and loans* of their balance sheets. If we try to account for the increased value of the deposit franchise, the insolvency often disappears. If deposits lose their stickiness, however, then the value of the deposit franchise will fall, and fear of insolvency can cause actual insolvency. For Silicon Valley Bank, that is exactly what happened next.

### **Silicon Valley Bank and the Panic of 2023**

Most of the mark-to-market losses shown in the bottom panel of Figure 1 had already manifested by the end of the third quarter of 2022. The existence of these losses was public. Everybody could see—without doing much work—that by

this measure Silicon Valley Bank was “insolvent” on paper. So why didn’t anything happen until March 2023?

The crucial shock arrived on March 8, 2023, when Silicon Valley Bank announced that it had sold \$24 billion of book value securities for a loss of \$1.8 billion, along with a plan to raise \$2.25 billion of new equity. This announcement contained new information. By selling the securities, SVB was now required to recognize these losses on its income statement and balance sheet—the losses were now “realized.” Banks would prefer not to realize losses, and it was reasonable to conclude that the bank did not have better options. Outside onlookers could ask reasonable questions: Maybe the deposit outflow is more severe than we thought? Maybe SVB will need to recognize even more losses to the point that they are actually insolvent? When these kinds of questions become salient, that is all it takes for uninsured deposits to lose their no-questions-asked property.

Once the no-questions-asked property is lost, the cost-benefit calculation for uninsured depositors changes. While history suggests that uninsured depositors have a good chance of ultimately getting all of their money back, it also suggests the possibility of significant delays before this happens. Thus, uninsured depositors knew that if Silicon Valley Bank failed, the potential costs of delay could be material, easily surpassing the transactions costs of switching banks. Many depositors seemed to share this thinking on Thursday, March 9, 2023. The \$42 billion withdrawn that day represented almost one-quarter of SVB deposits. Once the news of those withdrawals spread, even more depositors started asking questions. On Friday, March 10, the pace of withdrawals was so rapid that the authorities did not expect SVB to make it through the day.

The Federal Deposit Insurance Corporation took the bank into receivership in the morning—the first intraday receivership in its history. Its announcement on Friday, March 10, 2023, showed that uninsured depositors were right to be concerned: “All insured depositors will have full access to their insured deposits no later than Monday morning, March 13, 2023. The FDIC will pay uninsured depositors an advance dividend within the next week. Uninsured depositors will receive a receivership certificate for the remaining amount of their uninsured funds. As the FDIC sells the assets of Silicon Valley Bank, future dividend payments may be made to uninsured depositors” (FDIC 2023a).

On that Friday, uncertainty was all the Federal Deposit Insurance Corporation could promise for uninsured depositors. But the failure of Silicon Valley Bank, along with the murky outlook for uninsured depositors, led to a violation of the no-questions-asked property at many other banks. In particular, depositors ran from banks that had the same dangerous combination as SVB: a high percentage of uninsured deposits combined with significant unrealized losses on their assets. Two large banks were particularly hard-hit by deposit outflows that same Friday, March 10: Signature Bank, with \$110 billion in assets, had about \$10 billion in outflows, and First Republic Bank, with \$213 billion in assets, had about \$25 billion in outflows. Signature was particularly unprepared for this event. Many banks have collateral “pre-positioned” at the Federal Reserve, which involves filling out paperwork which designates certain bank assets and assures that the Fed would have a top priority

claim on those assets if the bank wants to use them as collateral for emergency discount-window borrowing. However, Signature Bank had no pre-positioned collateral, and thus would not have been able to access emergency borrowing from the Fed. This unpreparedness, combined with long-standing regulatory concerns about the management of the bank, convinced regulators to close the bank on Sunday, March 12, and create another FDIC receivership.

As the Federal Deposit Insurance Corporation sought to reassure the uninsured depositors, it faced a substantial obstacle. In any bank resolution, the FDIC is required by law to cover the insured depositors at the “least cost” to the deposit insurance fund. In many cases, another bank is often willing to take on all of the deposits of the failed bank in order to obtain those banking relationships for itself. However, the Silicon Valley Bank failure happened so fast that there was no time to arrange for such a purchase, nor any way to be certain that any such purchase would eventually occur. The FDIC would have had a hard time making a convincing case it was “least cost” to make noncontractual payments to uninsured depositors. The only way around the least-cost requirement is for the FDIC to invoke a “systemic risk exception,” enacted in 1991 legislation. This exception cannot be invoked casually: doing so requires a positive vote from the boards of both the FDIC and Federal Reserve, as well as approval by the US Secretary of the Treasury. Invoking the exception also brings close scrutiny to the FDIC, especially since Silicon Valley Bank had not previously been classified as being systemically important.

However, the Federal Deposit Insurance Corporation invoked the systemic risk exception on Sunday, March 12, and promised to protect uninsured depositors, thus showing how much can happen in two days (FDIC 2023b):

After receiving a recommendation from the boards of the FDIC and the Federal Reserve, and consulting with the President, Secretary Yellen approved actions enabling the FDIC to complete its resolution of Silicon Valley Bank, Santa Clara, California, in a manner that fully protects all depositors. Depositors will have access to all of their money starting Monday, March 13. No losses associated with the resolution of Silicon Valley Bank will be borne by the taxpayer. We are also announcing a similar systemic risk exception for Signature Bank, New York, New York, which was closed today by its state chartering authority. All depositors of this institution will be made whole. As with the resolution of Silicon Valley Bank, no losses will be borne by the taxpayer.

What changed from Friday, March 10, to Sunday, March 12? Given the information now publicly available, I speculate that on Friday, March 10, when Silicon Valley Bank was closed, the authorities had hoped that its failure was idiosyncratic and that the panic would not spread to other banks—or at least would not spread quickly. However, by Sunday, March 12, it was clear that the panic was contagious, and that the no-questions-asked property had been lost for unsecured deposits at banks with large unrealized losses.

The Federal Reserve also made an important policy announcement on that same Sunday, March 12, by introducing the Bank Term Funding Program, which broke new ground in emergency lending. Historically, the Federal Reserve would use the market value of securities as their collateral value and only allow banks to borrow up to that market value minus a small cushion. For the Bank Term Funding Program, it raised this collateral value to par for government securities. Thus, five-year duration government bonds issued in 2021, which were then trading at about \$0.85 on the dollar, would still count as a full \$1 of collateral from the discount window. Had this program been in place on March 9, and had Silicon Valley Bank possessed the operational capability to pledge all of its securities to the Fed, it would have been able to survive a much larger run.

The Bank Term Funding Program represented a major break from past Fed practice and required an invocation of the “unusual and exigent circumstances” clause from Section 13(3) of the Federal Reserve Act. This emergency authority, granted by Congress in 1935, went unused for more than 70 years before getting its first workout during the global financial crisis of 2008–2009. Since then, the Fed has decided that the “unusual and exigent circumstances” applied twice more, first during COVID-19 and again in the Panic of 2023.

The combination of the Federal Deposit Insurance Corporation using the systemic risk exception to reassure the uninsured depositors that they would be protected and the Federal Reserve using the unusual and exigent circumstances rule to create the Bank Term Funding Program succeeded in slowing down the outflow of deposits at many banks, but not in stopping it. First Republic Bank, which had seemed on Friday, March 10, only hours away from failure, was able to limp on for seven more weeks before being closed on May 1. This delay provided the FDIC enough time to organize an orderly auction, and the “least-cost” winning bid—from J. P. Morgan—included full assumption of all the uninsured deposits. Thus, the FDIC did not need to use the systemic-risk exception for the resolution of First Republic.

The government actions in March 2023 did not solve the underlying solvency concerns in the banking system. Interest rates remained high, so that the mark-to-market losses arising from higher interest rates remained. What the policies did achieve, however, was to reinforce the no-questions-asked property at a large number of troubled banks, thus allowing the deposit franchise of those banks to retain value. The franchise value from deposits then adds to the solvency strength of the bank, which reinforces the no-questions-asked property again, which further reinforces solvency strength, which further reinforces the no-questions-asked property—and then it is turtles all the way down.

## **The Regulation and Supervision of Silicon Valley Bank**

Bank oversight by the government consists of two main components: regulation and supervision. “Regulation” consists of the specific rules banks must follow; these rules are derived from federal and state law, and go through a detailed and

time-consuming administrative process for any changes. Overall, regulation moves slowly. “Supervision” is the day-to-day enforcement of these rules, with some limited discretion given to supervisors as circumscribed by the rules. In principle, supervision can be fast. In practice, it often is not.

Most bank regulation focuses on either capital or liquidity, but the connections between the two receive far less attention. For capital, banks face a variety of regulations, with standards harmonized internationally through the Basel Committee on Banking Supervision. While the specific rules are complex, they are all variations on a theme: set a minimum ratio where the numerator is some measure of capital and the denominator is some measure of “risk-adjusted” assets. In the simplest case, capital = equity and risk-adjusted assets = total assets, yielding a capital rule based on the equity-to-assets ratio. Indeed, preserving that simple equity-to-assets ratio has proved to be the binding constraint for many of the largest banks since the global financial crisis of 2008–2009. More complex versions of capital rules allow certain types of long-term debt into the numerator of the ratio and reduce the weight of relatively safe assets in the denominator. There are many good references that explain the details of these capital rules; for present purposes, the basic equity-to-assets ratio will suffice. At 7 percent, the ratio at Silicon Valley Bank at the end of 2019 was below the 9.66 percent average for insured banks in the United States, but still above the acceptable minimum of 4 percent.<sup>2</sup>

For liquidity, the classic form of regulation is “reserve requirements,” where a bank is required to maintain some minimum cash-to-deposits ratio, with the cash held either in its vault or at the central bank. As with the capital ratio, modern forms of liquidity regulation start with this simple formula and then allow more complex calculations in the numerator and denominator of the ratio.<sup>3</sup> But the main idea remains the same: the bank should have enough liquidity to meet even an unusual level of deposit withdrawals. However, these requirements are not meant to satisfy a full run of all depositors—no bank will have sufficient reserves to do that.

To enforce these capital and liquidity rules, government agencies employ supervisors to perform regular examinations. For Silicon Valley Bank, three agencies shared supervisory oversight. The Department of Financial Protection and Innovation was the chartering agency on behalf of the State of California. Because SVB was a member of the Federal Reserve System, the state-level department shared the regular examination work with the Federal Reserve Bank of San Francisco. The Fed also had authority over the bank holding company, SVB Financial. In this case, 98 percent of the assets of SVB Financial were held in the bank subsidiary, “Silicon Valley Bank” itself, which is the focus of our attention in this paper (Silicon Valley

<sup>2</sup>The historical development of capital rules is described in Haubrich (2020). The changes to international standards imposed after the global financial crisis are described in McNamara, Wedow, and Metrick (2019). A detailed discussion of the current rules and acceptable minimums in the United States is Scott and Labonte (2023). Aggregate statistics for insured banks at the end of 2019 is in FDIC (2020).

<sup>3</sup>A summary of the current international standards for liquidity regulation is given in Bank for International Settlements (2013). The specific application of these standards for Silicon Valley Bank is analyzed in Feldberg (2023a).

Bank Financial Group 2023). Because the San Francisco Fed has a much larger staff than the California Department of Financial Protection and Innovation, and also much more experience in the examination of large banks, the Fed took the lead on most of the examinations. Finally, as the insurer, the Federal Deposit Insurance Corporation is the “backup federal regulator,” and had access to all information from the supervisory process.

In the aftermath of the Panic of 2023, the performance of these regulations—and the supervisors that enforce them—came under close scrutiny, with several government agencies producing reports (DFPI 2023; Federal Reserve Board 2023; FDIC 2023d; Government Accountability Office 2023). The most comprehensive of these reports came from the Federal Reserve, the agency with the most supervisory resources dedicated to Silicon Valley Bank. The report concludes that SVB was in compliance with the specific capital and liquidity rules in place at the time of its failure, but faults itself for a set of recent reforms that kept SVB (and other similarly-sized banks) from having more stringent standards. They conclude, however, that there is no guarantee that these more stringent standards would have prevented SVB’s collapse. The Fed report also criticizes its own lack of supervisory zeal, saying that supervisors were aware of weaknesses at SVB, but were not aggressive enough in using their discretion to insist on timely corrections. This conclusion is well supported by evidence, but the remedy is not obvious. “Supervisors should do their jobs better” is not an easy goal to achieve.

To illustrate the challenge faced by supervisors, it is helpful to examine the largest source of concern—interest-rate risk. Many of the official reports emphasized the mismanagement of interest-rate risk by Silicon Valley Bank and the failure of regulation and supervision to correct this mismanagement. The Fed report concludes that “we need to evaluate how we supervise and regulate a bank’s management of interest rate risk. While interest rate risk is a core risk of banking that is not new to banks or supervisors, SVB did not appropriately manage its interest rate risk, and supervisors did not force the bank to fix these issues quickly enough” (Federal Reserve Board 2023, p. 3).

But on this point, the regulations themselves do not provide clear guidance to supervisors. In the computation of regulatory capital measures, most mark-to-market interest-rate losses are exempt from inclusion, allowing banks to ignore such losses in meeting required ratios. Even the regulatory “stress tests” performed on the largest banks, with binding implications for their dividend and capital policy, do *not* include a specific stress for interest-rate risk. I think it is asking too much of supervisors to enforce an ethos that is not already in the plain language of the rules.

In contrast, the European Union carefully monitors “interest-rate risk in the banking book” and uses a variety of mechanisms to discourage banks from taking on too much interest-rate risk. Their methods work: EU banks have lower interest-rate risk than do US banks. But these rules do not make the risk disappear—they just shift it to a different place. For example, one main mechanism in the EU financial system to move interest-rate risk away from banks is for more borrowers to pay

variable interest rates. In that case, if interest rates go up, bank balance sheets do not take a hit, but borrowers have to make higher payments. There, the risk of rising interest rates sits with borrowers.<sup>4</sup>

As long as actual investments take time to mature, the provision of liquidity will be a risky endeavor. That risk needs to sit somewhere, and no amount of supervisory vigilance can make this risk disappear. The European Union has chosen to shift this risk to borrowers, while the United States allows much more of that risk to sit on the balance sheets of banks. To the extent that the deposit franchise is a natural hedge for this interest-rate risk, there is a benefit in leaving this risk with banks. But there is also a potential cost, as we saw in March 2023.

## **Conclusion**

The alchemy of banking is that solvency and liquidity reinforce each other. Bank deposits remain money-like because of the assurance of solvency. Money-like deposits tend to be sticky within a given financial institution, because it is costly for depositors to switch banks and replicate an earlier long-term relationship with a new institution. This stickiness gives banks a form of monopoly power, the rents earned from this power comprise a large portion of bank profits, and thus the capitalization of these rents comprises a large portion of the bank's market value. Therein lies the alchemy. As long as the deposit base is stable, the expectation of future profits from deposit spreads adds to the solvency strength that supports the very same deposit stickiness. An information event that calls into question a bank's solvency can quickly become self-reinforcing, as a large component of the solvency disappears when the questions are asked openly. This is the main story of Silicon Valley Bank.

If the stability of the deposit base plays such a central role, why not simply increase the maximum level of deposit insurance? For example, during the global financial crisis of 2008–2009, the Federal Deposit Insurance Commission introduced the Transaction Account Guarantee (TAG) program, which provided unlimited insurance for non-interest-bearing accounts. This program was temporary—expiring at the end of 2010—and has generally been considered to be successful at calming depositors during the worst times of that crisis.

There are several challenges for making a program like the Transaction Account Guarantee permanent. First, when the program was deployed in 2009–2010, interest rates were headed toward zero. Depositors gave up very little by sticking with non-interest-bearing accounts in that environment. In the higher-interest-rate environment that we have today—and which we should expect to recur at least occasionally—the incentives for depositors to leave banks for other vehicles would be significantly stronger. Second, if we went further than the Transaction Account Guarantee and allowed unlimited deposit insurance even for interest-bearing

<sup>4</sup>For a discussion of the US approach to interest-rate risk as compared with the rest of the world, see Feldberg (2023b). For an analysis of EU interest-rate risk, see Dries et al. (2022).

deposits, we should not be surprised to see clever bankers finding ways to sell all kinds of financial products that are ultimately backed by an infinite safety net. For an intermediate solution of raising the deposit limit above \$250,000 but less than infinity, the same two problems exist, just in intermediate form.

The challenge of finding the “right” level of deposit insurance is a good example of the general problem of finding the right balance in financial regulation writ large. We should—of course!—try to do the best job we can in enforcing the regulations currently on the books. But in writing new regulations, we must recognize the tradeoffs. In general, bank regulation cannot change the total amount of risk in the economy, but it is highly effective at shifting that risk around. Banks, their customers, their substitutes (“shadow banks”), government, and taxpayers—what is our choice for who should hold what share and what kind of risk?

■ *I thank Greg Feldberg, Eric Hurst, Steven Kelly, Nina Pavcnik, Timothy Taylor, Rosalind Wiggins, and Heidi Williams for helpful comments and Vincent Arnold, Anmol Makhija, and Carey Mott for excellent research assistance.*

## References

- Bank for International Settlements.** 2013. *Basel III: The Liquidity Coverage Ratio and Liquidity Risk Monitoring Tools*. Basel: Bank for International Settlements.
- California Department of Financial Protection and Innovation (DFPI).** 2023. “Review of DFPI’s Oversight and Regulation of Silicon Valley Bank.” Sacramento: DFPI.
- Dries, Jonathan, Benjamin Klaus, Francesca Lenoci, and Cosimo Pancaro.** 2022. “Interest Rate Risk Exposures and Hedging of Euro Area Banks’ Banking Books.” In *Financial Stability Review, May 2022*, 67–70. Frankfurt: European Central Bank.
- Drechsler, Itamar, Alexi Savov, and Philipp Schnabl.** 2021. “Banking on Deposits: Maturity Transformation without Interest Rate Risk.” *Journal of Finance* 76 (3): 1091–1143.
- FDIC.** 2020. “Quarterly Banking Profile: Fourth Quarter 2019.” *FDIC Quarterly* 14 (1).
- FDIC.** 2023a. “FDIC Creates a Deposit Insurance National Bank of Santa Clara to Protect Insured Depositors of Silicon Valley Bank, Santa Clara, California.” PR-16-2023. <https://www.fdic.gov/news/press-releases/2023/pr23016.html>.
- FDIC.** 2023b. “Joint Statement by the Department of the Treasury, Federal Reserve, and FDIC.” PR-17-2023. <https://www.fdic.gov/news/press-releases/2023/pr23017.html>.
- FDIC.** 2023c. *FDIC’s Supervision of Signature Bank*. Washington, DC: FDIC.
- FDIC.** 2023d. *Options for Deposit Insurance Reform*. Washington, DC: FDIC.
- Federal Reserve Board.** 2023. Review of the Federal Reserve’s Supervision and Regulation of Silicon Valley Bank. Washington, DC: Federal Reserve Board.
- Feldberg, Greg.** 2023a. “Lessons from Applying the Liquidity Coverage Ratio to Silicon Valley Bank.” Yale School of Management, March 27. <https://som.yale.edu/story/2023/lessons-applying-liquidity-coverage-ratio-silicon-valley-bank>.
- Feldberg, Greg.** 2023b. “US Banks’ Interest-Rate Risk Reporting and Regulation: A Comparative Context.” Yale School of Management, April 26. <https://som.yale.edu/story/2023/us-banks-interest-rate-risk-reporting-and-regulation-comparative-context>.



- Flannery, Mark J., and Sorin M. Sorescu.** 2023. "Partial Effects of Fed Tightening on U.S. Banks' Capital." Unpublished.
- Government Accountability Office.** 2023. "Bank Regulation: Preliminary Review of Agency Actions Related to March 2023 Bank Failures." GAO-23-106834. <https://www.gao.gov/products/gao-23-106834>.
- Haubrich, Joseph G.** 2020. "A Brief History of Bank Capital Requirements in the United States." Federal Reserve Bank of Cleveland Economic Commentary 2020-05.
- Haubrich, Joseph G., and Tristan Young.** 2019. "Trends in the Noninterest Income of Banks." Federal Reserve Bank of Cleveland Economic Commentary 2019-14.
- Holmstrom, Bengt.** 2015. "Understanding the Role of Debt in the Financial System." BIS Working Paper 479.
- Jiang, Erica Xuewei, Gregor Matvos, Tomasz Piskorski, and Amit Seru.** 2023. "Monetary Tightening and U.S. Bank Fragility in 2023: Mark-to-Market Losses and Uninsured Depositor Runs?" NBER Working Paper 31048.
- McNamara, Christian, Michael Wedow, and Andrew Metrick.** 2019. "Basel III B: Basel III Overview." *Journal of Financial Crises* 1 (4): 59–69.
- Scott, Andrew P., and Marc Labonte.** 2023. *Bank Capital Requirements: A Primer and Policy Issues*. CRS Report R47447. Washington, DC: Congressional Research Service.
- Silicon Valley Bank (SVB).** 2019. "Call Report at the Close of Business December 31, 2019." RSSD-ID 802866. <https://cdr.ffiec.gov/public/ManageFacsimiles.aspx#>.
- Silicon Valley Bank (SVB).** 2020a. "Q4 and FY 2019 Financial Highlights." [https://s201.q4cdn.com/589201576/files/doc\\_financials/2019/q4/Q4'19-IR-Presentation\\_1.31.20.pdf](https://s201.q4cdn.com/589201576/files/doc_financials/2019/q4/Q4'19-IR-Presentation_1.31.20.pdf).
- Silicon Valley Bank (SVB).** 2020b. "Call Report at the Close of Business December 31, 2020." RSSD-ID 802866. <https://cdr.ffiec.gov/public/ManageFacsimiles.aspx#>.
- Silicon Valley Bank (SVB).** 2021. "Call Report at the Close of Business December 31, 2021." RSSD-ID 802866. <https://cdr.ffiec.gov/public/ManageFacsimiles.aspx#>.
- Silicon Valley Bank (SVB).** 2022. "Call Report at the Close of Business December 31, 2022." RSSD-ID 802866. <https://cdr.ffiec.gov/public/ManageFacsimiles.aspx#>.
- Silicon Valley Bank Financial Group.** 2023. "Form 10-K for the Fiscal Year Ended December 31, 2022." Washington, DC: United State Securities and Exchange Commission <https://d18rn0p25nwr6d.cloudfront.net/CIK-0000719739/f36fc4d7-9459-41d7-9e3d-2c468971b386.pdf>.
- US Department of the Treasury.** 2023. "Daily Treasury Par Yield Curve Rates." [https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily\\_treasury\\_yield\\_curve&field\\_tdr\\_date\\_value\\_month=202311](https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_yield_curve&field_tdr_date_value_month=202311).



## Is Pay Transparency Good?

Zoë Cullen

**T**he well-known case of Lilly Ledbetter suggests that information about the pay of coworkers can empower low earners to stand up to pay discrimination. In 1998, Ledbetter received a covert message from a male colleague sharing his salary and alerting her to the differences in the paychecks that she had received compared to 15 other all-male “area managers” at the same company, with similar experience and doing the same work. She used this information to press legal charges against her employer for unfair compensation. Her case became the basis for the 2009 Lilly Ledbetter Fair Pay Act, which removed the statute of limitation for pay discrimination lawsuits (Phillips 2009).

It is tempting to infer from this case that greater transparency in salaries would help more disadvantaged workers to negotiate for higher pay. That reasoning has supported the passage of many pay transparency mandates (European Commission 2017; Obama 2014). The public discussion around Ledbetter’s high-profile case, however, reflects narrow thinking about the impact of pay transparency. Such reasoning ignores how employers respond to greater transparency by changing their hiring and pay strategies. It also ignores outcomes beyond pay gaps, like productivity and wage levels. Plus, the case focuses on only one kind of pay transparency, transparency between peers doing the same work. But what happens when the pay of the boss or boss’s boss becomes transparent, when pay information circulates beyond a single business to workers in competing businesses, or when the hiring manager who sets the wage learns about other wages being offered to a candidate?

■ *Zoë Cullen is Assistant Professor of Business Administration, Harvard Business School, Boston, Massachusetts. Her email address is [zcullen@hbs.edu](mailto:zcullen@hbs.edu).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.1.153>.

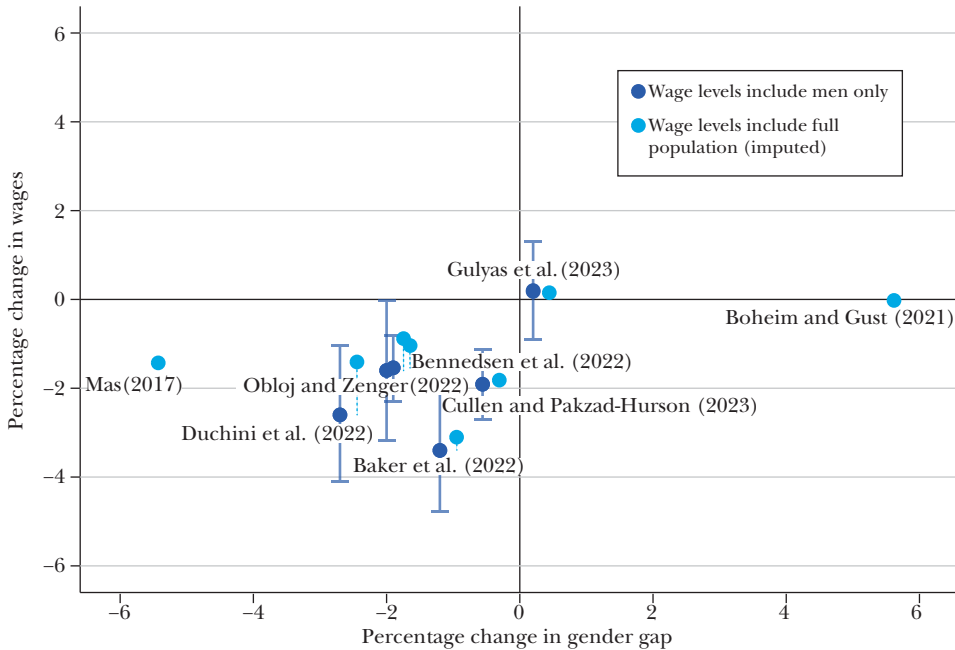
To understand the economics behind pay transparency, it is useful to categorize the types of pay transparency into three buckets: horizontal pay transparency, where coworkers at the same organization are informed of each other's pay (the Lilly Ledbetter case); vertical transparency, where transparency extends to different layers of seniority within an organization; and cross-firm transparency, where workers and/or employers have access to the pay information of competing firms and organizations.

Horizontal pay transparency gained traction politically because it could hold a bad actor accountable for "unfair" pay differences, as in the Lilly Ledbetter case. In specific situations where roles are clearly defined, this kind of pay transparency could be a starting point to identify discriminatory gaps in pay that are prohibited by the Equal Pay Act of 1963 and related laws. Research has shown that employers have responded to horizontal pay transparency by achieving equality through lower average pay overall. Figure 1 highlights this key trade-off that policymakers face when implementing horizontal pay transparency laws revealing coworkers pay gaps. Each data point represents the findings from the evaluation of a pay transparency mandate.<sup>1</sup> In the cases where transparency achieved greater pay equalization between men and women—those in the lower left quadrant of the graph—the reduction in pay gap was accompanied by an overall reduction in wages. Economic theory offers an explanation. Horizontal pay transparency between coworkers within a firm created spillovers between negotiations; specifically, a \$1 raise for one worker became more costly due to renegotiations with other workers who have the expectation of equal pay, causing employers to bargain more aggressively with each worker. Moreover, when wages were not equalized under horizontal transparency, research has shown that workers paid visibly less than some of their peers (which can be the majority of workers) felt disgruntled and exerted less effort.

In contrast, vertical pay transparency and cross-firm pay transparency, while less equipped to hold specific organizations accountable for discrimination, have proven capable of raising productivity and raising wages by reducing information frictions in the labor market. Vertical pay transparency increases workers' information about what they could earn if they were to be promoted. Because employees typically underestimate the steepness of financial rewards from promotion, vertical transparency raises expectations about potential earnings and has proven to boost effort and productivity in meritocratic environments. Cross-firm pay transparency, achieved through salary benchmarks like Glassdoor or salary ranges in job posts, informs prospective candidates about which employers pay more than others and leads applicants, especially those underpaid, to redirect their search toward higher paying firms and more favorable pay negotiations. Cross-firm pay transparency policies have also informed firms what their competitors are paying, increasing competition and putting upward pressure on wages. These pay transparency policies shine the light outward, away from coworkers under the same employer, toward vertical and cross-firm pay differences.

<sup>1</sup>For details behind Figure 1, see the discussion later in this paper under the section heading *A Framework for Effects of Horizontal Pay*, which included Table 2 with a full list of the studies in this figure.

*Figure 1*  
**Effect of Pay Transparency on Wages, Existing Studies**



*Source:* Each observation is a study of a pay transparency policy, with study details and results describe further in Table 2.

*Note:* The y-axis is the percentage change in wages, and the x-axis represents the percentage change in the gender wage gap. Dark blue dots represent the percentage change in wages for men only, with standard error bars, while light blue dots represent the percentage change in wages for all workers.

Information frictions in the labor market run deep. Not only are some workers in the dark about the pay of coworkers, but they are often in the dark about what pay they can expect in the future and the pay they could earn if they switch jobs or invest in training. Employers also face barriers to learning about market wages. The reasons for these information frictions are myriad—taboos around salary discussions, obstacles to credible communication, laws guarding against collusion, preferences for privacy, and strategic obfuscation. While many of the sources for information frictions are hard to overcome entirely, we do know that pay transparency policies make a difference in people’s perceptions about pay and, in turn, change behavior. The wage-setting games that employers engage in, and the career choices that workers make, respond to pay transparency tools and pay transparency policies.

The pay transparency lever is, in the end, stunningly cheap and powerful. However, this comes with a warning. Among the lessons learned in the study of pay transparency is that more information is not always better. Thus, we describe an open field for pay transparency studies to identify where greater and more equal access to information can improve economic outcomes. We begin this paper with a

review of the different kinds of pay transparency policies around the world. We then turn to the implications of horizontal pay transparency, revealing pay gaps between coworkers. Then, we discuss how alternative kinds of pay transparency, including vertical and cross-firm pay transparency, can affect labor market outcomes. We touch upon nonlabor market outcomes affected by pay transparency, including political preferences, and then conclude.

## What Drives Variation in Wages across Workers?

To understand how pay transparency affects labor market outcomes, we must begin by understanding why people might be paid differently, if at all. In the United States, we observe significant variation in pay. Take for instance the sample of 600,000+ large and medium-sized firms that use Automatic Data Processing, Inc. (ADP) to process their payroll, allowing us to inspect differences in paychecks: for any two employees chosen at random within one of these companies, there is an average gap of 60 percent in their paychecks. If we take two people in the same position title (from a list of 10,000 standardized titles) without regard to their employer, the average gap is 46.7 percent. If we restrict it to two people who share the same position title at the same company, this gap is 10.7 percent.<sup>2</sup>

Some of the variation in compensation can come from differences in *how* people are compensated. In principle, total compensation includes benefits, bonuses, equity, in-kind transfers, and amenities, in addition to hourly or salary pay. However, most datasets lack the level of detail necessary to observe all components of compensation, and, as a consequence, research and policies are designed around what is observable. According to several in-depth studies of compensation, base pay (hourly wage or annual salary) empirically comprises more than 80 percent of financial compensation across all jobs and 100 percent of financial compensation for most jobs (Barkume 2004; Bryson and Freeman 2010; Gittleman and Pierce 2013). While we do not have standard measures of nonfinancial compensation, such as benefits and workplace amenities, we typically assume that base pay is the most important form of compensation for most employees.

Some of the differences in pay can be explained by differences in how people perform in their positions—some people will have more experience or exert more effort, and employers may reward and encourage high performance with higher pay. For simplicity, we will refer to these types of pay differences as “justified” as long as pay would be equal once we measured pay per unit of output. Some of the variation in pay comes from differences across employers. For example, some organizations are more productive than others, so individuals with similar skills are able to generate more revenue in some businesses than others—perhaps the equipment and training are higher quality—and the productive firms may want to attract

<sup>2</sup>Original calculations by the author, using ADP payroll data from 2022.

more workers by offering higher pay. Generally, the perception is that pay differences stemming from production differences across firms are also “justified” in the sense that performance-adjusted pay may still be equal.<sup>3</sup> Finally, some of the differences in pay can be traced back to how workers and employers negotiate pay. Some workers may be less assertive and some employers may be more aggressive in their negotiations. Some employers may share more of their profits with workers, and some employers may express discriminatory views by making lower wage offers to some groups of their employees. In light of our discussion about pay transparency and the policy objective to achieve “fair pay,” we will refer to all pay differentials that cannot be explained by differential performance as “unjustified,” keeping in mind that a special class of unjustified pay pertains to discrimination against protected classes of workers. In the next section, we discuss how this categorization of pay differences is viewed in the eyes of policymakers.

## **Pay Transparency Policies around the World**

While pay transparency policies pertain broadly to visibility across positions and firms, in practice the legal and political discussion has focused on pay differences between people working the same job, at a similar performance level, and under the same employer, narrowing in on so-called “unjustified” pay differences. In the United States, this focus on horizontal pay differences arose due to transparency policies designed to bring to light violations of the Equal Pay Act (EPA) of 1963, which prohibited wage discrimination based on sex, and related laws further prohibiting wage discrimination on the bases of race, color, national origin, religion, age, marital status, political affiliation, and disability. A violation requires pay differences to exist between workers in jobs that require “substantially equal skill, effort, and responsibility and be performed under similar working conditions within the same establishment.”

In the United States, the most popular form of pay transparency policy protects workers’ ability to discuss pay amongst themselves. Nationwide, the National Labor Relations Act of 1935 protects workers’ rights to discuss pay with their coworkers, but this legislation, designed to support collective bargaining, did not stipulate punishments for employers who retaliated against employees that disclosed or inquired about pay. Twenty states have since enacted laws that explicitly penalize employers for retaliating against any employee who discusses pay with their coworkers. Additionally, four states require that the employer either disclose pay range information or pay statistics about coworkers upon request, including information about benefits.

Table 1 describes a (probably partial) count of 18 types of pay transparency policies in 32 countries around the world, illustrating a popular aim of bringing to light pay discrimination through horizontal pay transparency. Canada’s Pay Equity Act uses prototypical language: “[T]he purpose of this Act is to achieve pay equity

<sup>3</sup>Of course, to the extent that some workers are simply lucky to secure jobs at more productive firms, we may question how “justified” these pay differences are.

Table 1

**Partial List of Government Pay Transparency Mandates**

Policy	Places implemented		Type
	Countries	Localities	
Reporting Gender Wage Gap Statistics by Occupational Group	Denmark (2006), Austria (2011), Belgium (2012), Lithuania (2017), Iceland (2018), Spain (2021), Canada (2020), Israel (2022)	NJ* (2018)	Horizontal
Transparent Pay Criteria for Job	Portugal, Peru (2017), Croatia		Horizontal
Ad Hoc Gender Pay Audits	Greece (2011), Turkey (2003), Ireland (2014), the Netherlands, Costa Rica		Horizontal
Right of Workers to Talk: Protected Disclosure & Inquiry	United States* (2014)	CA (2016), CO (2012), CT (2015), DE (2016), DC (2015), IL (2004), LA* (2013), ME (2009), MD (2016), MA (2018), MI, MN (2014), NE (2019), NV (2017), NH (2015), NJ (2013), NY (2016), OR (2016), VT (2013), WA (2018)	Horizontal
Wages or Wage Range Disclosure Employees Upon Request	Germany (2017), Chile	NV (2021), CT (2021), WA, RI (2023), CA (2023) for	Horizontal
Collective Bargaining Information Sharing	United States (1935)		Horizontal
Reporting Gender Wage Gap Statistic	Australia (2013), UK (2018), France (2018), Switzerland (2020), Italy (2021), Japan (2023)		Horizontal, Vertical
Public Salaries	Norway (1814), Sweden (1903), Finland (1800s), Iceland		Horizontal, Vertical, Cross-Firm
Government Employee Public Salaries	Estonia (2012)	CA (2010), CT (2015), DC (2014), IL (2021), LA (2023), MN (2011), NH (2009), NJ (2010), OR (2016), VA (2016), VT (2017), Córdoba, Argentina (2017)	Horizontal, Vertical, Cross-Firm
Expected Wages in Job Postings	Austria (2011), Slovakia (2018), Latvia (2019), Lithuania (2019)		Cross-Firm
Wage Ranges in Job Postings		Ontario, Canada (2019), CO (2021), CA (2023), WA (2023), and four US cities	Cross-Firm
Wage Ranges for Job Applicants Upon Request		CA (2018), MD (2020), WA, NV (2021), CT (2021), RI (2023), and two US cities	Cross-Firm
Public Labor Cost Index	Slovakia (2019)		Cross-Firm
Antitrust Safety Zone Permitting Salary Benchmarking	United States (1996)		Cross-Firm
Ban on Salary Discussion Among HR Professionals	United States (2016)		Cross-Firm
Prohibition of Salary History Bans		MI (2018), WI (2018)	Cross-Firm
Salary History Ban		AL (2019), CA (2018), CO (2021), CT (2019), DE (2017), (2019), IL (2019), ME (2019), MD (2020), MA (2018), NV (2021), NJ (2018), NY (2017), OR (2017), PR (2017), RI (2023), VT (2018), WA (2019), NC* (2019), PA* (2018), VA* (2019), Ontario, Canada (2019), and 20 US cities	Cross-Firm

Source: HR Dive (2022a); NLWC (2020); International Labour Organization (2022); Women's Bureau (2022); Frey (2021); European Commission (2017); HR Dive (2022b); Dreisbach (2014); Obama (2014); Perez-Truglia (2020); NJDOL (2018); City of Toledo (2019); City of Cincinnati (2019); Behrmann and Osborn (2018); DOJ and FTC (2016); Gobierno de la Provincia de Córdoba (2022); Johanson (2021); Green, Ceron, and Tanzi (2022); Midorikawa (2022); Nadworny (2022); Percivalle (2021); Valdez and Polar (2018); Shepherd (2022); Frimmel et al. (2022); Bennedsen, Larsen, and Wei (2022); Matthews (2019); Barry (2018); Gely and Bierman (2003); O'Donoghue (2023); Connecticut Office of the State Comptroller (2015); District of Columbia (2014); Illinois State Comptroller (2021); Minnesota Management and Budget (2011); State of New Hampshire (2009); State of New Jersey Transparency Center (2010); State of Oregon (2016); Commonwealth of Virginia (2016); and State of Vermont (2017).

Note: States or countries marked with \* indicate the policy only applies to state employees or federal contractors, indicates laws passed but not yet effective.



through proactive means by redressing the systemic gender-based discrimination in the compensation practices” (Government of Canada 2018). Among the 27 countries that have implemented laws with the intent of addressing the gender pay gap, 24 are covered by policies that reveal pay gaps between coworkers. An additional three countries (Portugal, Croatia, and Peru) require organizations to make transparent the pay criteria for a job and the reason for any pay differences, and two countries (Germany and Chile) require that the employer disclose pay information about coworkers upon request. In Germany, for example, an employee at a firm with more than 200 employees can ask their employer for pay statistics about their occupation, broken down by gender. Each of these forms of transparency promote awareness of horizontal pay differences in the hopes of holding firms more accountable for justifying pay differences.

Other motivations for pay transparency policies include cultivating public trust and accountability. In Norway, Sweden, and Finland, income tax records are public. Norway’s tax authority explains they make the records public because “the opportunity to check the tax assessment process in general, as well as for individuals and groups of taxpayers, must be available in our society.”<sup>4</sup> In Estonia, as well as in 18 US states, government employee salaries are made public, ostensibly to enhance government accountability. The United States also has national policies aimed at promoting competition between firms, including a ban on salary discussions among competing firms and state bans that protect the previous salary history of job candidates.

Enforcing pay transparency policies boils down to which forms of compensation can be tracked. While the objective of most policies is to enforce fair overall compensation, with language inclusive of benefits and transfers beyond base pay, in practice most enforcement plans hinge on limited data collected through the government agencies, such as the Equal Employment Opportunity Commission (EEOC) in the United States, or privately collected by employees and assessed by the court system. To the extent sources lack detailed information about benefits, equity, and nonpecuniary compensation, these components of compensation will be less tightly enforced. Similarly, research on the effectiveness of these policies also rests on what compensation is documented either in survey or administrative data. Hence, in what follows, our evidence on the impact of pay transparency focuses on what appears in payroll data and household or establishment surveys conducted by governments, and we will miss shifts toward hard-to-observe components of compensation.

## **Horizontal Pay Transparency: Revealing Coworker Pay Gaps**

Horizontal pay transparency policies, designed to identify unjustified pay gaps and pin them to a particular employer, only have a chance of changing behavior if they are shifting perceptions about coworker pay.

<sup>4</sup> The quotation is from the website of the Skatteetaten, the Norwegian Tax Administration, at <https://www.skatteetaten.no/en/forms/search-the-tax-lists/>.

On the one hand, we typically think that coworkers gossip with each other regularly and would talk about any topic that is important to them, including pay. On the other hand, we may doubt that pay is an easy topic of conversation. As the journalist Margaret Littman (2001, p. 39) put it: “You may know about your colleague’s sex life, your friend’s drinking problem, or what your neighbor really thinks of her mother-in-law. But you probably don’t know what they take home in each paycheck.” Indeed, surveys documenting beliefs about coworker salaries suggest significant misperceptions (Sun et al. 2021; Lawler 1965; Cullen and Perez-Truglia 2023).

The likely reasons that coworkers do not know about each other’s pay include workplace policies that promote pay secrecy and the disinclination of many workers to publicize their own pay. In a survey of full-time US workers in 2018, Sun, Rosenfeld, and Denice (2021) find that 35.4 percent of workers report pay discussion is discouraged in their workplace. Social scientists have pointed to social norms and taboos that curtail conversations about salaries, hypothesizing that such information touches on feelings about one’s personal worth to society (Edwards 2005; Lawler 1965; Trachtman 1999). Cullen and Perez-Truglia (2023) find quantitative support for social concerns around asking and sharing salary information: at the commercial bank that they study, 89 percent of respondents report they would feel uncomfortable if they had to ask a coworker about their salary, and 80 percent would forgo cash rewards to prevent their employer from sharing their individual earnings with coworkers. Willingness to disclose salary information appears to decline with age (Goldfarb and Tucker 2012) and relative standing, such that fears of resentment or competition may fuel the taboo more than status concerns about appearing well-off (Cullen and Perez-Truglia 2023).

Following pay transparency mandates, employees report greater access to information; for example, in US states that passed laws protecting workers’ rights to inquire and disclose pay between 2010 and 2018, the share of private sector workers reporting their employer prevents them from discussing pay fell from 33 percent to 10 percent during this window, while other states experienced a modest decline (Cullen and Pakzad-Hurson 2023; Sun, Rosenfeld, and Denice 2021). While this finding suggests that pay transparency policies shifted perceptions about what information sharing is permissible, more research is warranted to make a tight connection between specific pay transparency policies and knowledge about coworker pay.

In principle, pay transparency could also shift perceptions about one’s relative productivity. Suppose you interpret the revelation that your pay is lower than your coworker’s as a sign that your performance must also be lower. While we have limited direct evidence on this, in one setting, Cullen and Perez-Truglia (2022) test whether experimentally introducing pay transparency shifts employee perceptions about their performance rating relative to their peers in the same position and establishment. They find that employees started out with highly accurate perceptions about their relative performance ratings, and pay information had little effect on this. More studies are necessary before concluding this is broadly true.

### A Framework for Effects of Horizontal Pay Transparency

Theoretically, how should pay transparency policies that reveal pay gaps between coworkers affect labor market outcomes? For some intuition, consider a two-part scenario: a worker learns that a colleague with the same job is earning significantly more than she is. She reasons that her employer must be willing to pay a higher wage for the work she is doing, or she sees the opportunity for courts to enforce equal pay. For either reason, she can seek to renegotiate her wage. The first part of the story is that information will lead workers to take action, demanding higher pay as Lily Ledbetter did. We have reason to believe this would happen with regularity: in a study using data from Hired.com, an online labor platform, Roussille (2021) showed that revealing median pay to job applicants led them to request at least the median pay in their applications. We also have evidence that when lesser-paid employees learn about unjustified pay gaps, they express discontentment about pay and reduce effort, cooperation, and work hours if those pay differences go unaddressed (Breza, Kaur, and Shamdasani 2018; Card et al. 2012; Cullen and Perez-Truglia 2022; Ockenfels, Sliwka, and Werner 2015), as predicted by Akerloff and Yellen (1990).

But there is a second part to the story. How will employers hire and set pay when they anticipate transparency? Consider wage negotiations under full pay transparency. A worker knows the wages of her peers, but also recognizes that her wage will be visible to her coworkers. The employer can credibly reject her demand for a raise by saying, “If I give you a higher salary, I’ll have to give everyone else a raise too, and I can’t afford that.” Under pay secrecy, the worker might have been skeptical of such a claim and bargained aggressively regardless, but due to transparency, the worker grasps the (true and costly) ramifications of asking for more than her coworkers make. This dynamic is akin to “stiffening the backbone” discussed in other contexts (Kreps and Wilson 1982; Milgrom and Weber 1982). At the same time, workers will tend to bargain less aggressively over pay when being hired, because they are less willing to risk rejection during the initial negotiation given the high likelihood of learning the prevailing wages for coworkers and renegotiating down the road. This dynamic is akin to “freeriding off others” motivation discussed by Kuhn and Gu (1998, 1999).

For these reasons, full horizontal pay transparency leads to an unintended side effect: if workers all get the same wage and cannot negotiate this wage upward, the firm effectively gains the power to set the wage. To maximize its profit, the firm acts like a monopsonist and sets a relatively low wage.<sup>5</sup> In this setting, pay transparency increases the de facto bargaining power of the employer, becoming the enforcement

<sup>5</sup>It is worth noting the parallels between wage-setting under transparency and the literature on monopsony power. Under secrecy, employers may have leeway to pay the marginal hire more than other workers, effectively price discriminating, but under transparency, a raise for the marginal hire implies a raise for everyone who renegotiates with this information. The standard monopsony model in Alan Manning’s (2003) book *Monopsony in Motion: Imperfect Competition in Labor Markets* assumes employers must pay a single wage to homogenous workers, which leads to wage markdowns relative to the marginal product of workers. If firms could perfectly price discriminate, using the veil of pay secrecy, they would pay the marginal worker a wage equivalent to their marginal product. In this way, transparency may offer a microfoundation for monopsonistic wage-setting behavior.

mechanism for a low wage. For a formal model of a wage-setting game which conveys the economic forces present when transitioning from secrecy to transparency, and extensions to other bargaining environments, see Cullen and Pakzad-Hurson (2023).<sup>6</sup>

To get into the mindset of a company implementing transparency, consider Buffer, a company that chose to go fully transparent with its pay as a mid-sized start-up (Gallani et al. 2017a, b). First, it considered posting the names of employees, including the chief executive officer, and exactly what they earned. Before doing so, it reviewed compensation closely and made sure that steps had been taken to mitigate any unjustified pay differences between employees. After posting, employees still had questions about pay gaps and whether they should be paid more equally. Thus, the company went even further, and posted a salary calculator, so that any employee or job candidate could enter their role, responsibilities, experience, and location into the salary calculation and see the corresponding paycheck. In essence, when the company shifted from pay secrecy to pay transparency, it also adopted a formulaic pay structure. The new system would make it costly to make an exception for an employee—in fact, the whole formula would have to change along with the pay of those affected.

To assess theoretical predictions of the full equilibrium effect of pay transparency, we turn to the evaluations of large-scale pay transparency policies from five different countries: the United States, Canada, the United Kingdom, Austria, and Denmark. Results have been published in eight independent studies that track wages around the time the policy is enacted.<sup>7</sup> Because these policies have been focused largely on the goal of combatting pay discrimination, the studies have centered on whether these policies have reached their stated objective: closing wages gaps between men and women. Because our equilibrium model predicts that pay will be more equal, but lower, as a consequence of shifting bargaining power toward the firm, we pull from each paper information about both the wage gap and the wage levels. We report study details and results in Table 2.<sup>8</sup> In Figure 1 (presented

<sup>6</sup>In the context of the Cullen and Pakzad-Hurson (2023) model, under full pay secrecy, workers make a single take-it-or-leave-it offer. This maximizes expected worker surplus and expected wages (Williams 1987). Under full pay transparency, the reverse happens, the employer sets a posted price, akin to an employer take-it-or-leave-it offer. And, consequently, the employer maximizes their surplus. Increasing transparency shifts bargaining power toward the employer: the expected employer profit is strictly increasing in level of transparency, and expected wages are strictly decreasing with an increase in transparency.

<sup>7</sup>We select studies using the procedure from Cullen and Pakzad-Hurson (2023). First, the policy studied must be referred to as a “pay transparency” policy or a related term. Second, it must study the policy in the context of a real-world labor market. Third, it must assess the effect of the policy on the wages of all employees in that labor market.

<sup>8</sup>Five studies (in three countries) evaluate mandates to publish wage statistics by gender (Bennedsen et al. 2022; Böheim and Gust 2021; Duchini et al. 2022; Gulyas, Seitz, and Sinha 2023). UK law requires firm-level statistics be made public, while Austria and Denmark require averages by occupational group be disclosed to employees. Three studies evaluate policies that mandate posting of individual salaries in university or municipal contexts (Baker et al 2022; Mas 2016; Obloj and Zenger 2020). The final study in the table concerns “Right of Workers to Talk” laws, prohibiting US employers in states that have passed such legislation from punishing workers who internally discuss or inquire about salaries (Cullen and Pakzad-Hurson 2023).

*Table 2*  
**Eight Studies of Horizontal Pay Transparency Laws**

<i>Study</i>	<i>Baker et al.</i> (2022)	<i>Bennedsen et al.</i> (2020)	<i>Böheim and Gust</i> (2021)	<i>Duchini et al.</i> (2022)
Setting	Canadian universities	Danish private sector	Austrian private sector	UK private sector
Policy	Posting individual salaries	Disclosure of relative earnings by gender	Disclosure of relative earnings by gender	Disclosure of relative earnings by gender
Men's wage effect (standard error)	-0.034 (0.007)	-0.015 (0.0073)	0.005	-0.026 (0.008)
Women's wage effect (standard error)	-0.022 (0.006)	0.0036 (0.0043)	-0.008	0.003 (0.014)
Share men	0.725	0.7	0.42	0.53
W:M pay ratio (pre-policy)	0.89	0.84	0.78	0.82
Imputed overall wage effect	-0.031	-0.010	-0.000	-0.014
<i>Study</i>	<i>Gulyas, Seitz, and Sinha</i> (2023)	<i>Mas</i> (2016)	<i>Obloj and Zenger</i> (2022)	<i>Cullen and Pakzad-Hurson</i> (2023)
Setting	Austrian private sector	CA public sector	US universities	13 US states
Policy	Disclosure of relative earnings by gender	Posting individual salaries	Posting individual salaries	Right of workers to talk
Men's wage effect (standard error)	0.002 (0.004)	-0.014 (0.017)	-0.016 (0.008)	-0.019 (0.007)
Women's wage effect (standard error)	0.001 (0.004)	-0.07 (0.021)	0.005 (0.004)	-0.016 (0.005)
Share men	0.58	0.99	0.614	0.58
W:M pay ratio (pre-policy)	0.75	2.80	0.93	0.74
Imputed overall wage effect	0.002	-0.014	-0.009	-0.018

*Source:* Baker et al. (2022): Numbers drawn from Table 4 Col. 4, Table 2. Bennedsen et al. (2020): Numbers drawn from Table 3 Col. 7, Table 1. Duchini et al. (2022): Numbers drawn from Table 3 Col. 1, Table 2. Böheim and Gust (2021): Numbers drawn from Table 1, Table 4, Panel D. Row 2. Gulyas, Seitz, and Sinha (2023): Numbers drawn from Table 1, Table B2 Col. 2, Footnote 6. Mas (2016): Numbers drawn from Table 2 Col. 5 Row 3, Table 3 Col. 2 Row 3. Additional numbers drawn from the California municipal pay website at <https://publicpay.ca.gov/Reports> and Reese (2019). Obloj and Zenger (2022): Numbers drawn from Table 1 Col. 6, page 5. Cullen and Pakzad-Hurson (2023): Numbers drawn from Table C.1, Figure D.5.

earlier), and we plot the percent change in the gender wage gap (along the  $x$ -axis) against the percent change in overall wage levels (along the  $y$ -axis) for each study. Following Cullen and Pakzad-Hurson (2023), we include two data points from each study when available. The darker points capture the effect size directly reported in the paper and refers to the effect of pay transparency on men's wages along with the 95 percent confidence interval. The lighter points reflect the imputed estimate of transparency's effect on the overall population.<sup>9</sup>

Why did some pay transparency policies have no effect at all on wages? Our bargaining theory offers one explanation. Workers must start out with individual bargaining power for pay transparency to create spillovers between individual negotiations. In many labor markets, workers bargain under a collective agreement (Bhuller et al. 2022). Cullen and Pakzad-Hurson (2023) document that in environments where individuals have a high degree of bargaining power to begin with, namely in occupations with low unionization rates, pay transparency mandates were followed by significant wage reductions, while highly unionized occupations experienced negligible wage changes.

Our bargaining theory predicts lower and more equal pay when individual negotiations are newly anchored to one another. Could it also be that employers adjust pay in fear of demoralizing employees who observe unjustified pay gaps? In their study of Denmark's requirement that firms with more than 35 employees internally report wage statistics by gender and occupation, Bennedsen et al. (2020) find that average wages per employee in firms just above the policy size threshold fell by 2.8 percent relative to those just below the policy size threshold, and one year later, productivity (measured by average sales per employee) was 2.7 percent lower above the policy threshold. This study is our best evidence to date on the presence of a productivity response to inequality that may persist in equilibrium in spite of rebargaining. This fall in productivity is consistent with the demotivating effect of learning one earns less than their peers; it is also consistent with a reaction to slower wage growth overall. More research with granular measures of output is necessary to disentangle the two mechanisms.

## **Alternative Innovative Designs for Pay Transparency**

Overall, evidence on the effects of revealing wage gaps between coworkers has taught us about two unintended consequences when pay transparency is horizontal. First, the information spillovers between the negotiations of employees under the same employer can shift the de facto bargaining power toward the employer, lowering wages. Second, peer coworkers compare themselves to each other. In cases where employers fail to equalize wages, lesser-paid employees can experience lower morale and lower effort. In this section, we will look at how vertical and cross-firm

<sup>9</sup>These data points for Mas (2017) and Böheim and Gust (2021) are omitted because wage results for men are not reported. Further details are in the online Appendix.

pay transparency policies affect labor market outcomes. While negotiation spillovers and interpersonal comparisons might still be at play, these effects are muted. These alternative pay transparency policies operate by educating workers about the full range of opportunities to earn higher wages by applying to higher paying firms and reaching promotion. Cross-firm pay transparency policies also educate employers about market wages. Many of the pay differences revealed would be considered “justified” pay gaps.

To see the potential of these alternative pay transparency designs, we will walk through key moments where workers and employers form beliefs about the pay being offered in the marketplace and use this information to guide their decisions. In turn, it will elucidate where more information can potentially shift beliefs and behavior.

On the supply side of the labor market, let a worker’s earnings be a function of several choices: human capital investments, the intensity of job search, the effectiveness of their negotiations, and their effort on the job. Each of these choices are made under uncertainty. The pay of others can shape the beliefs about expected returns to each of these actions, and in turn influence their choices. For example, information about the value of a promotion might come from water cooler conversations about what others have earned, which in turn can influence motivation and effort. Job advertisements might lack pay information or list a wide range, so it may be the pay of a former coworker now at this firm that anchors pay expectations and determines whether to apply for such a job. Similarly, the pay of former graduates may be the best available information on the returns to a particular education or training program, and encourage or dissuade enrollment for that reason.

On the demand side of the labor market, employers make wage-setting and hiring decisions as a function of what they believe their competitors are offering, as well as what they believe their employees know about these outside offers. *Prima facie* evidence that employers face uncertainty about market pay is the prevalence of consulting firms that offer salary benchmarking services, including Abbott, Langer, and Associates; Korn Ferry; Mercer; Radford; and Willis Towers Watson.

Some aspects of misperceptions and uncertainty about pay have long been embedded in our canonical economic models of the labor market. Sixty years ago, Stigler (1962) formalized the idea that job seekers face limited information about the full set of outside offers available to them. The subsequent modern search and matching models retain this feature (Diamond 1982; Mortensen 1982, 2005; Pissarides 1985; Postel-Vinay and Robin 2002). This limited information that workers have about the pay at alternative employment opportunities can lead to wage differences between similar workers who face different perceptions of their opportunities. However, many forms of misperceptions and uncertainty about pay are not captured by our go-to models of the labor market. For example, economists rarely consider misperceptions on the employer side (Cullen, Li, and Perez-Truglia 2022).

Recent empirical work has made headway by documenting pay misperceptions and uncertainty with survey tools. With the aid of “information experiments”—in

which treatment groups and a control group are exposed to different information before making a decision—researchers have demonstrated how pay information shifts beliefs and behavior. In the following section, we describe documented misperceptions and their estimated effects, and combine these empirical facts with theoretical frameworks. We highlight instances further pay transparency studies could illuminate consequential information frictions and policy solutions.

### **Transparency in Returns to On-the-Job Effort: Vertical Pay Transparency**

A rich literature on optimal contracts indicates that incentives should rise steeply over one's career, precisely to encourage employees to work hard and stick with their employer (for an overview of the literature, see Lazear 2018; Dewatripont, Jewitt, and Tirole 1999; Gibbons and Waldman 1999b). What are people's actual perceptions of how steeply these incentives rise, and thus, how might pay transparency influence perception and behavior? In a classic study, Lawler (1965) surveyed 326 managers from four privately owned US companies and found respondents underestimated the salaries of those in higher positions. This pattern of underestimating vertical inequality has been replicated in other settings. For example, Cullen and Perez-Truglia (2022) carried out a guessing game at a commercial bank, where employees could earn a sizable financial prize if they guessed the average salary of their boss's position within  $\pm 5$  percent of the truth. Employees underestimated the truth on average by 14.1 percent.

When people systematically underestimate what their superiors earn, the news content of pay transparency inflates perceptions about overall inequality and positively impacts expected future earnings in environments where workers are upwardly mobile. Cullen and Perez-Truglia (2022) find that, with every 10 percent boost in perceived managers' salary, subjects projected that their own earnings would be 1.7 percent higher five years down the road. News of larger pay gaps with higher-ups did not generate resentment, though in theory it might have. While employees responded negatively to learning their peers earned more than expected, these social concerns did not extend to comparisons with higher-ups. Linking these survey data with administrative data on employee performance revealed that employees also increased their sales revenue by 1.1 percent, they sent 1.3 percent more emails, and worked 1.5 percent longer hours for each 10 percent upward shift in perception about managerial pay.<sup>10</sup> In the context of public sector workers in Sierra Leone, Deserrano, Kastrau, and León-Ciliotta (2021) showed the effect of vertical transparency differed depending on whether the environment was perceived as meritocratic. In meritocratic segments of the Ministry of Health, perceptions of steeper salary raises upon promotion increased effort, measured by the number of home health visits. In nonmeritocratic segments, perceptions of steeper pay raises lowered morale and reduced home health visits because learning about higher superior pay did not raise own earnings expectations.

<sup>10</sup>In a similar vein, Flynn (2022) finds that National Hockey League players shifted their efforts towards more highly compensated strategies like offense, following pay transparency.



What does theory predict happens to wages when vertical pay gaps are made transparent? Like horizontal pay transparency, vertical pay transparency creates a link between the negotiations of employees bargaining with a single employer. On the one hand, employers will recognize that a raise to someone's boss in turn inflates what their employees will ask for upon promotion, which provides an incentive for employers to bargain more aggressively. On the other hand, the motivating effects of revealing a steeper-than-expected reward structure in the organization could very well dominate the effect of shifting bargaining power, and the net effect on wages could be positive or neutral.

Some pay transparency policies, existing or under consideration, can be considered vertical pay transparency. Most notably, a 2020 UK policy requires UK-listed companies with more than 250 employees to report the ratio of pay between the median, 25th percentile, and 75th percentile employees and the chief executive officer (Clark 2019). In the United States, the Securities and Exchange Commission enacted in 2017 a regulation requiring reporting of the ratio of the median worker to the pay of the chief executive officer (US Securities and Exchange Commission 2015). Of course, these ratios might not help workers to learn about potential gains in the next step of one's own personal earnings trajectory in any precise way, though they could still be positive news about expected pay. Another set of policies that reveal vertical pay are internal pay grids, which have been implemented in many public sector settings world-wide (examples in Table 1) and allow employees to observe the earnings trajectory within their organization. Currently, we lack direct evidence on the causal effects of implementing these policies.

### **Transparency in Returns to Job Search**

Pay is only one feature of the job, but it may be an especially influential one. Belot, Kircher, and Muller (2019) show that the wage information in job postings affects where people apply. As noted earlier, economists have long observed that job search is a process in which workers can only reasonably collect information about a tiny fraction of suitable jobs. Using data on online US job market listings from Burning Glass, Arnold, Quach, and Taska (2022) find that only 30–40 percent of job postings have information about the pay of the job directly listed. According to Hall and Krueger (2012), only 23 percent of recent hires had a clear expectation about how much the job paid at the time they first interviewed. Without upfront pay information, do workers misdirect their applications? What do they assume about pay in the absence of explicit information from the employer?

An empirical regularity across a range of settings is that workers anchor on their own wage (or most recent wage) when forming beliefs about the pay of job opportunities, whether it be internal promotion opportunities, like the employees from a multibillion-dollar corporation in Southeast Asia studied in Cullen and Perez-Truglia (2022), or the potential outside wage offers reported by a representative sample of workers in Denmark (Hvidberg, Kreiner, and Stantcheva 2020). One implication is that low-wage workers, in particular, underestimate the pay associated with job opportunities. Jäger et al. (2021) showed this by sampling from the

German workforce and asking employees hypothetical questions about the wage they would receive if they left their current job and accepted another job within three months. When these subjective beliefs about wages were compared with the actual wages of observationally identical coworkers who switched employers, they proved to be systematically lower than reality among low-earners and higher than reality among high-earners. The beliefs of the unemployed are an exception to this pattern: their job prospects can dip below their previous wages, and hence anchoring on previous wages can lead to over-optimism (for example, as argued in Arni 2013; Krueger and Mueller 2016; Mueller, Spinnewijn, and Topa 2021; Spinnewijn 2015).

Theoretically, more information about the pay at various job opportunities would lead candidates to direct their applications toward higher-paying firms, all else equal. Relatively low earners would be the most motivated by the news to renegotiate their pay or apply elsewhere. Such increase in competition generated by directed job search places upward pressure on wages, and pushes toward pay compression as underpaid workers relocate. (Employers may also learn more about the pay at competing firms, a channel we discuss in the next subsection.)

Both observational and experimental studies capture the positive impact that cross-firm pay transparency can have for employees and job seekers. Using Danish administrative employer-employee data, Caldwell and Harmon (2019) offer evidence that news about other job offers arrives through the networks of former coworkers, and those informed are able to use the information to negotiate raises and switch to higher paying jobs. In a natural experiment on Hired.com, a job-matching platform geared toward engineers, Roussille (2021) showed that when women were informed about the median offers that other candidates received *across* employers, the information resulted in higher offer salaries for women on the platform by 2.6 percent, fully closing the gender gap.

New laws requiring salary information to be included explicitly in the advertisement for a job have taken effect in a few US states and cities (Colorado, New York City, Washington, and California) as well as a few European countries (Austria, Slovakia, Latvia, and Lithuania). The implementation of these policies may allow researchers to study what happens in equilibrium when cross-firm pay transparency policies are mandated broadly. In unpublished surveys results, along with Roussille and Jaeger, we interviewed over 2,000 employers using ZipRecruiter, a nation-wide US job posting platform, and asked their expectations about the law's full impact. Employers generally expect wages to either rise (33.1 percent) or stay the same (65.1 percent). Only 1.7 percent expect wages to fall. On average, they expect wages to rise 2.4 percent. The majority (56.5 percent) also expect salary ranges to increase the quality of their applicants. While close to 50 percent of employers believe turnover will not change, 35.0 percent believe their higher-performing employees will stay longer and 42.7 percent expect higher churn among their lower-performing employees.

Evidence from other settings on the effects of these laws suggests that these expectations of US employers may be on target. Slovakia's cross-firm pay

transparency law that went into effect on May 1, 2018, required firms nationwide to include an expected salary in all job advertisements. Skoda (2022) tracked job applications and final wages in Slovakia before and after the law. Before the passage of the law, the share of job postings with salary information ranged from 10 percent in some job titles to 60 percent in others. After the reform, nearly all job titles listed expected pay in the job advertisement. Consequentially, job applicants in Slovakia applied to a more diverse set of opportunities, spanning a greater number of sectors and wider array of job titles. The earnings of those hired after the reform were, on average, 3 percent higher than the wages of those hired before the reform. A similar law in Austria required postings to include a minimum wage offer (Frimmel et al. 2022), and similar law in Colorado required wage ranges in job postings (Arnold, Quach, and Taska 2022); both found an increase in wages posted after the reform, consistent with effects documented in Slovakia. Researchers are now awaiting data on the effects in New York City, California, and Washington, which enacted similar legislation in late 2022 and early 2023.

### **Employer-Side Pay Transparency**

Most economic models suggest workers would be forthcoming about their outside options in order to secure higher pay from their current employer, but empirically, such renegotiations are far from guaranteed. Dube, Naidu, and Reich (2022) offer job opportunities to Walmart workers and find that higher-paid outside offers prompted quits to rise faster than renegotiations. Anecdotally, employers think about the possibility of workers getting poached when setting initial salaries. In their human resources textbook, Berger and Berger (2008, p. 125) write, “No organization wants to waste their financial resources by paying too high relative to the market; and those who pay too low risk unwanted turnover from employees looking for a better deal elsewhere”—a view that suggests employers may not expect a chance to match the outside offer if and when it arrives. Based on their survey of 1,350 human resources professionals who report setting pay for new and current employees, Cullen, Li, Perez-Truglia (2022) find that 81 percent report limited or no access to outside offer information of their employees, and 20 percent report not even having easy access to their own internal pay records, which some organizations consider sufficiently sensitive to restrict managers’ access.

Firms also face legal barriers to collecting and retaining some pay information. For example, firms are legally prohibited from discussing salaries directly with their competitors, a law intended to prevent collusion (DOJ and FTC 2016). In addition, some speculate that firms face constraints storing data that can be used against them in discrimination lawsuits (Adler 2022).

How would more information about competitor pay, or outside options of job candidates, affect wages and hiring? Using tools from auction theory, one can generalize competitive labor market models so that each firm is uncertain about the distribution of market wages, even in equilibrium. In such models, benchmarks that reveal the wage distribution lead to compression. However, under

standard technical conditions, such benchmarks raise the mean wage by sharpening competition between firms. This result proceeds from the linkage principle (Milgrom and Weber 1982); the information revealed by a benchmark more strongly links each firm's payment to the aggregate market conditions, which erodes that firm's information rents.

When firms do not have perfect information about the value of a candidate, they may infer additional information through the bids of other firms. In other words, they may learn more about what a candidate could produce by seeing other firm wage offers. In this case, a "common values" auction model (that is, an auction in which the item will have the same value to all bidders, but bidders have different information about that value) can shed light on what will happen in equilibrium (Krishna 2009). Students of this literature will be familiar with the "winner's curse": if firms all bid their (noisy) signal about the value of the candidate, the winning bid is likely to be one of the parties with the highest positive measurement error. For fear of being that party, and thus falling victim to the winners' curse, firms will shade down their bids. When more information about competing bids becomes public knowledge, concerns about a winners' curse subside and firms feel less need to shade down their bids. Again, bids and final wages rise in the presence of greater pay transparency.

Cross-firm pay transparency policies typically provide pay information to prospective employees and also to competing firms. In their survey of employers about US state legislation mandating employers include salary ranges in their job postings, Cullen, Jaeger, and Roussille found that over half of employers expressed interest in the salary range information included in job advertisements similar to theirs, and 33.6 percent reported that they will use this information when setting their own salary ranges.<sup>11</sup>

Finding a persuasive empirical strategy to disentangle the impact that pay information has through firm beliefs and employee beliefs poses challenges, especially in contexts where the information is made simultaneously available to both sides of the market. In one setting studied, however, market wage data was made only available to employers. Cullen, Li, and Perez-Truglia (2022) study the roll-out of a proprietary salary benchmark that allowed clients of the largest payroll processing firm to have access to detailed salary information, based on their full database of payroll for 650,000 US employers. For employers who gained access to this high-quality salary benchmark, pay-setting converged toward the median pay in the marketplace, in this way reducing wage inequality (measured as the absolute dispersion from the median) in new hire salaries by over 25 percent. Among low-skilled positions (capturing approximately half of new hires) where the salary data was likely particularly informative, dispersion fell by 40 percent. Wage levels rose between 1–2 percent overall, and wages for new hires in lower-skill groups rose by

<sup>11</sup> The results refer to an unpublished survey hosted on ZipRecruiter, available to employers at the point they post a job advertisement on this nation-wide internet job board: 1,630 employer responses were collected.

over 6 percent along with corresponding boosts in retention (16 percent higher among low-skilled positions). Overall, the evidence suggests information about market pay affected firms' beliefs and behavior, reducing inequality in pay within a position title, raising wages among low-skilled jobs, and improving efficiency through longer-lasting job matches.

Much more research on this topic is necessary to comprehend the extent of information frictions on the firm side more fully and to design policies to improve market efficiency. For example, are firms only learning about the outside options of workers, or are they also learning about the value an employee can bring to their firm? As many of the policies that have potential to improve directed worker search will also affect employer beliefs about competitor pricing, what is the interaction of these two channels?

### **Transparency in Returns to Education**

Policies related to pay transparency might also seek to affect the decisions that workers or employers make well before a salary negotiation. For example, do young people have an accurate perception of how their education choices are related to earnings in the short and long run? How might pay transparency affect these perceptions and choices?

In an early attempt to address this question, Betts (1996) surveyed undergraduates at the University of California at San Diego and asked them to guess the average national earnings for students, both for recent graduates and also later in their lives. While the questions were about national averages, not expectations about personal returns to education (the relevant theoretical object), answers could be compared to government earnings statistics to measure accuracy of beliefs. The students significantly underestimated the slope of the wage profile, but the error around the average earnings of recent college graduates was modest (5.8 percent was the mean percentage error). Knowledge was not equal across types of students: students from poorer families exhibited much larger errors estimating the salaries of college graduates.

More recent studies have also suggested that additional pay transparency related to education choices can alter aspirations. In a representative sample of US heads of households, Bleemer and Zafar (2018) find that information about returns to schooling increased intention to attend college by 0.2 standard deviations over baseline expectations among a representative US sample of households. They also document that disadvantaged households have bigger errors in their perceptions of returns to a college degree, and that giving the information on returns to education helps close the socioeconomic gap in schooling expectations. Similarly, Wiswall and Zafar (2014) show information to students at New York University about the earnings and labor supply by degree and major. The students updated beliefs about their own earnings expectations when shown information about earnings in the population. Moreover, the average odds of shifting to the two-highest paying majors—economics/business and engineering/computer science—relative to the humanities shifted by 46 percentage points and 72 percentage points, respectively.

In a parallel study of students conducted in a low-income country, the Dominican Republic, Jensen (2010) noted that over 70 percent reported relying on the people they knew in their community for information about earnings—in part, presumably, because they lacked other sources of information on pay. Eighth-graders facing the decision to enter secondary school systematically underestimated the returns to a secondary education, undershooting by 14 percent on average. When eighth-grade boys received information about what a typical 30 or 40 year old earns on average as a function of education level, six months later these participants could recollect this information, and over the next four years, they ended up completing more years of education (between 0.20 and 0.35 more years).

The current pay transparency mandates, as captured earlier in Table 1, are not tailored to shed light on returns to education. Moreover, publicizing the returns to education has not been a primary motivation for pay transparency policies. The US policy most related to this goal is a website designed and maintained by researchers at the US Census Bureau (2020), the Post-Secondary Employment Outcomes (PSEO) data, where students can look up information about the salaries negotiated by recent graduates as a function of their college and major. One potential shortcoming of making pay information available through government websites is that search behavior is imperfect, and people with lower literacy have proven less likely to find relevant information (Fuster et al. 2022). As is the case with most forms of pay transparency, the information is “within reach” with only a few conversations, but mandates and clever policies are required to make the information readily accessible at the point decisions are made.

### **Nonlabor Market Outcomes Affected by Pay Transparency: Tax Delinquency, Happiness, and Preferences for Redistribution**

Pay transparency policies might affect outcomes outside the labor market, including tax compliance, happiness, and preferences for redistribution. Some of the literature draws directly on pay transparency legislation; some investigates the effects of when information about pay or income is provided by the researcher in a field experiment.

In a few countries, pay transparency seeks to curb tax delinquency by making income public information. For example, Norway, Finland, Sweden, and Iceland publicize individual incomes “to check the tax assessment process in general” (in the words of the Skatteetaten, the Norwegian tax authority). Bø, Slemrod, and Thoresen (2015) look at what happened to tax compliance in Norway around the time that the tax records became searchable online in 2001 and find that business owners reported higher income, rising by about 3 percent on average.

Moreover, publicizing tax returns and income across the economy had the unintended side effect of correcting misperceptions about the income distribution. A number of studies have documented misperceptions about pay by asking questions about the overall income distribution. In much the way workers anchor on their

own wage when asked about the pay of peers, respondents anchor to their own local environment when asked about inequality across the economy. Because inequality in one's local environment is more muted on average than across the economy, people systematically underestimate economy-wide inequality. Additional information about income or pay shifts perceptions on average toward greater inequality (Hauser and Norton 2017; Hvidberg, Kreiner, and Stantcheva 2020; Kuziemko et al. 2015; Norton and Ariely 2011).

These more accurate perceptions of the distribution of income and pay may have implications for happiness and for preferences over redistribution. The information contained in pay transparency about relative standing could be a significant driver of overall well-being, given that happiness hinges to some comparison to others. Indeed, Luttmer (2005) used self-reported happiness data, along with individual and local-level income from US data on Public Use Microdata Areas, and found that one's happiness declined as one's neighbor's income rose, holding own income constant. Perez-Truglia (2020) showed pay transparency increased the happiness gap between the rich and poor: in Norway, when tax records became highly visible online, people learned about their relative positions: those with lower incomes experienced a drop in happiness, while those with high incomes experienced a symmetric increase in happiness.

Several studies in which researchers provided accurate information about pay or income suggest that demand for redistribution could also respond to pay transparency. In a study in Buenos Aires, Argentina, with 1,100 representative households, Cruces, Perez-Truglia, and Tetaz (2013) exposed participants to accurate information about their own position in the income distribution and found that people who mistakenly believed they were middle class reacted by increasing their support for government welfare. Similarly, Kuziemko et al. (2015) randomized information treatments to a large internet panel of Americans, conveying to people the true US income distribution as well as growth in inequality since 1980. Indeed, the truth shifted people's beliefs about inequality and led them to express concerns about inequality. However, the authors concluded that the new information only increased demand for redistribution when subjects viewed government as effective at combating inequality. For those that believed alternative methods were more effective solutions (for example, hard work), learning the true extent of inequality did little to change demand for redistribution. In the context of a financial institution in Southeast Asia, Cullen and Perez-Truglia (2022) find that, when employees learn their peers earn more than they were expecting, they report higher dissatisfaction with the extent of inequality at the firm.

## **Conclusion**

Horizontal pay transparency policies that reveal pay gaps between coworkers doing similar work at the same firm characterize many pay transparency policies that have been put in place over the past two decades, designed to bring to light

unjustified pay gaps inside an organization. Recent research has revealed that these policies reduce pay gaps but also have unintended spillovers between worker negotiations that lower worker bargaining power and wages. In contrast, vertical and cross-firm pay transparency policies that ameliorate information frictions in the labor market more broadly have shown potential to improve motivation and talent allocation and sharpen competition, and, in so doing, raise wages, productivity, and equity. These alternative pay transparency policies are not designed to draw attention to employers who pay different wages to similar workers, but instead to educate workers about the full range of opportunities to earn higher wages when they make decisions about what type of work to pursue, how hard to work, and for whom they work. Pay transparency policies can also educate employers about market wages, with procompetitive effects.

■ *Thank you to Simon Jaeger, Shengwu Li, Bobak Pakzad-Hurson, Ricardo Perez-Truglia, and Nina Roussille for early feedback and invaluable intellectual contributions to the research agenda. Special thanks to Julia Gilman for her research assistance.*

## References

- Adler, Laura.** 2022. “From the Job’s Worth to the Person’s Price: Pay-Setting, Gender Inequality, and the Changing Understanding of Fair Play.” Unpublished.
- Akerlof, George A., and Janet L. Yellen.** 1990. “The Fair Wage-Effort Hypothesis and Unemployment.” *Quarterly Journal of Economics* 105 (2): 255–83.
- Arni, Patrick.** 2013. “What’s in the Blackbox? The Effect of Labor Market Policy on Search Behavior and Beliefs: A Field Experiment.” Unpublished.
- Arnold, David, Simon Quach, and Bledi Taska.** 2022. “The Impact of Pay Transparency in Job Postings on the Labor Market.” Unpublished.
- Baker, Michael, Yosh Halberstam, Kory Kroft, Alexandre Mas, and Derek Messacar.** 2022. “Pay Transparency and the Gender Gap.” *American Economic Journal: Applied Economics* 15 (2): 157–83.
- Barkume, Anthony J.** 2004. “Using Incentive Pay and Providing Pay Supplements in U.S. Job Markets.” *Industrial Relations: A Journal of Economy and Society* 43 (3): 618–33.
- Barry, Ellen.** 2018. “Happy ‘National Jealousy Day’! Finland Bares Its Citizens’ Taxes.” *New York Times*, November 1. <https://www.nytimes.com/2018/11/01/world/europe/finland-national-jealousy-day.html>.
- Behrmann, Anna, and Kathy Osborn.** 2018. “In Effort to Increase Employees’ Bargaining Power, U.S. Senators Petition Federal Antitrust Agencies to Revisit Benchmarking Guidelines.” *JD Supra*, July 13. <https://www.jdsupra.com/legalnews/in-effort-to-increase-employees-26757>.
- Belot, Michéle, Philipp Kircher, and Paul Muller.** 2019. “Providing Advice to Jobseekers at Low Cost: An Experimental Study on Online Advice.” *Review of Economic Studies* 86 (4): 1411–47.
- Bennedsen, Morten, Birthe Larsen, and Jiayi Wei.** 2022. “Wage Transparency and the Gender Pay Gap: A Survey.” Copenhagen Business School Department of Economics Working Paper 17-2022.
- Bennedsen, Morten, Elena Simintzi, Margarita Tsoutsoura, and Daniel Wolfenzon.** 2022. “Do Firms Respond to Gender Pay Gap Transparency?” *Journal of Finance* 77 (4): 2051–91.



- Berger, Lance A., and Dorothy R. Berger.** 2008. *The Compensation Handbook*. New York: McGraw-Hill.
- Betts, Julian R.** 1996. "What Do Students Know about Wages? Evidence from a Survey of Undergraduates." *Journal of Human Resources* 31 (1): 27–56.
- Bhuller, Manudeep, Karl Ove Moene, Magne Mogstad, and Ola L. Vestad.** 2022. "Facts and Fantasies about Wage Setting and Collective Bargaining." *Journal of Economic Perspectives* 36 (4): 29–52.
- Bleemer, Zachary, and Basit Zafar.** 2018. "Intended College Attendance: Evidence from an Experiment on College Returns and Costs." *Journal of Public Economics* 157: 184–211.
- Bø, Erlend E., Joel Slemrod, and Thor O. Thoresen.** 2015. "Taxes on the Internet: Deterrence Effects of Public Disclosure." *American Economic Journal: Economic Policy* 7 (1): 36–62.
- Böheim, René and Sarah Gust.** 2021. "The Austrian Pay Transparency Law and the Gender Wage Gap." IZA Discussion Paper 14206.
- Breza, Emily, Supreet Kaur, and Yogita Shamdasani.** 2018. "The Morale Effects of Pay Inequality." *Quarterly Journal of Economics* 133 (2): 611–63.
- Brütt, Katharina, and Huaiping Yuan.** 2022. "Pitfalls of Pay Transparency: Evidence from the Lab and the Field." Tinbergen Institute Discussion Paper TI 2022-055/I.
- Bryson, Alex, and Richard Freeman.** 2010. "How Does Shared Capitalism Affect Economic Performance in the United Kingdom?" In *Shared Capitalism at Work: Employee Ownership, Profit and Gain Sharing, and Broad-based Stock Options*, edited by Douglas L. Kruse, Richard B. Freeman, and Joseph R Blasi. Chicago: University of Chicago Press.
- Caldwell, Sydnee, and Nikolaj Harmon.** 2019. "Outside Options, Bargaining, and Wages: Evidence from Coworker Networks." Unpublished.
- Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez.** 2012. "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction." *American Economic Review* 102 (6): 2981–3003.
- City of Cincinnati.** 2019. An Ordinance No. 83. <https://www.cincinnati-oh.gov/cityofcincinnati/equity-in-cincinnati/city-of-cincinnati-s-salary-equity-ordinance/>.
- City of Toledo.** 2019. Enacting a New Toledo Municipal Code Chapter 768, Pay Equity Act to Prohibit the Inquiry and Use of Salary History in Hiring Practices in the City of Toledo. O-173-19. <https://toledo.legistar.com/LegislationDetail.aspx?ID=3911922&GUID=F109B51B-1960-4700-AF21-B45D77BAF464&Options=&Search=>.
- Clark, Greg.** 2019. "New Executive Pay Transparency Measures Come into Force." *Gov.UK*, January 1. <https://www.gov.uk/government/news/new-executive-pay-transparency-measures-come-into-force>.
- Commonwealth of Virginia.** 2016. "Virginia State Salaries." DataPoint. <https://www.datapoint.apa.virginia.gov/salaries.php> (accessed on November 7, 2023).
- Connecticut Office of the State Comptroller.** 2015. "Payroll Summary 2015." OpenPayroll. [https://openpayroll.ct.gov/#/year/2015/card/payroll\\_summary](https://openpayroll.ct.gov/#/year/2015/card/payroll_summary) (accessed on November 7, 2023).
- Cruces, Guillermo, Ricardo Perez-Truglia, and Martin Tetaz.** 2013. "Biased Perceptions of Income Distribution and Preferences for Redistribution: Evidence from a Survey Experiment." *Journal of Public Economics* 98: 100–112.
- Cullen, Zoë, Shengwu Li, and Ricardo Perez-Truglia.** 2022. "What's My Employee Worth? The Effects of Salary Benchmarking." NBER Working Paper 30570.
- Cullen, Zoë B., and Bobak Pakzad-Hurson.** 2023. "Equilibrium Effects of Pay Transparency." *Econometrica* 91 (3): 765–802.
- Cullen, Zoë, and Ricardo Perez-Truglia.** 2022. "How Much Does Your Boss Make? The Effects of Salary Comparisons." *Journal of Political Economy* 130 (3): 766–822.
- Cullen, Zoë, and Ricardo Perez-Truglia.** 2023. "The Salary Taboo: Privacy Norms and the Diffusion of Information." *Journal of Public Economics* 222: 104890.
- Deloitte.** 2021. *Sustainable Finance Disclosure Regulation: Is the Financial Industry Ready for the Big One?* London, UK: Deloitte.
- Department of Justice (DOJ), and Federal Trade Commission (FTC).** 2016. *Antitrust Guidance for Human Resource Professionals*. Washington, DC: DOJ and FTC.
- Deserrano, Erika, Philipp Kastra, and Gianmarco León-Ciliotta.** "Promotions and Productivity: The Role of Meritocracy and Pay Progression in the Public Sector." CEPR STEG Working Paper 037.
- Dewatripont, Mathias, Ian Jewitt, and Jean Tirole.** 1999. "The Economics of Career Concerns, Part I: Comparing Information Structures." *Review of Economic Studies* 66 (1): 183–98.
- Diamond, Peter A.** 1982. "Wage Determination and Efficiency in Search Equilibrium." *Review of Economic Studies* 49 (2): 217–27.

- District of Columbia.** 2014. "Public Employee Salary Information." Department of Human Resources, Government of the District of Columbia. <https://dchr.dc.gov/public-employee-salary-information> (accessed on November 7, 2023).
- Dreisbach, Tom.** 2014. "Pay Secrecy Policies at Work: Often Illegal, and Misunderstood." *NPR*, April 13. <https://www.npr.org/2014/04/13/301989789/pay-secrecy-policies-at-work-often-illegal-and-misunderstood>.
- Dube, Arindrajit, Suresh Naidu, and Adam D. Reich.** 2022. "Power and Dignity in the Low-Wage Labor Market: Theory and Evidence from Wal-Mart Workers." NBER Working Paper 30441.
- Duchini, Emma, Stefania Simion, Arthur Turrell, and Jack Blundell.** 2022. "Pay Transparency and Gender Equality." Unpublished.
- Edwards, Matthew A.** 2005. "The Law and Social Norms of Pay Secrecy." *Berkeley Journal of Employment and Labor Law* 26: 41–63.
- European Commission.** 2017. *Pay Transparency in the EU: A Legal Analysis of the Situation in the EU Member States, Iceland, Liechtenstein and Norway*. Brussels: European Commission.
- Feldstein, Martin, and James Poterba.** 1984. "Unemployment Insurance and Reservation Wages." *Journal of Public Economics* 23 (1-2): 141–67.
- Flynn, James.** 2022. "Salary Disclosure and Individual Effort: Evidence from the National Hockey League." *Journal of Economic Behavior and Organization* 202: 471–97.
- Frey, Valerie.** 2021. "Can Pay Transparency Tools Close the Gender Wage Gap?" *OECD*. <https://www.oecd.org/gender/Pay-Transparency-2021-Policy-Brief.pdf>.
- Frimmel, Wolfgang, Bernhard Schmidpeter, Rene Wiesinger, and Rudolf Winter-Ebmer.** 2022. "Mandatory Wage Posting, Bargaining and the Gender Wage Gap." Johannes Kepler University of Linz, Department of Economics Working Paper 2202.
- Fuster, Andreas, Ricardo Perez-Truglia, Mirko Wiederholt, and Basit Zafar.** 2022. "Expectations with Endogenous Information Acquisition: An Experimental Investigation." *Review of Economics and Statistics* 104 (5): 1059–78.
- Gallani, Susanna, Tiffany Y. Chang, Brian J. Hall, and Jee Eun Shin.** 2017a. "Buffer.com." Harvard Business School Case 917-019.
- Gallani, Susanna, Tiffany Y. Chang, Brian J. Hall, and Jee Eun Shin.** 2017b. "Buffer.com (B)." Harvard Business School Case 917-020.
- Gely, Rafael, and Leonard Bierman.** 2003. "Pay Secrecy/Confidentiality Rules and the National Labor Relations Act." *Journal of Labor and Employment Law* 6: 120–56.
- Gibbons, Robert, and Kevin J. Murphy.** 1992. "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence." *Journal of Political Economy* 100 (3): 468–505.
- Gibbons, Robert, and Michael Waldman.** 1999a. "A Theory of Wage and Promotion Dynamics inside Firms." *Quarterly Journal of Economics* 114 (4): 1321–58.
- Gibbons, Robert, and Michael Waldman.** 1999b. "Careers in Organizations: Theory and Evidence." In *Handbook of Labor Economics*, Vol. 3B, edited by Orley C. Ashenfelter and David Card, 2373–2437. Amsterdam: North Holland.
- Gittleman, Maury, and Brooks Pierce.** 2013. "How Prevalent is Performance-Related Pay in the United States? Current Incidence and Recent Trends." *National Institute Economic Review* 226 (1): R4–R16.
- Gobierno de la Provincia de Córdoba.** 2022. *Portal de transparencia*. <http://transparencia.cba.gov.ar/>.
- Goldfarb, Avi, and Catherine Tucker.** 2012. "Shifts in Privacy Concerns." *American Economic Review* 102 (3): 349–53.
- Government of Canada.** 2018. *Pay Equity Act*. Ottawa: Government of Canada Publications.
- Government of UK.** 2019. "The Companies (Directors' Remuneration Policy and Directors' Remuneration Report) Regulations 2019." <https://www.legislation.gov.uk/ukxi/2019/970/contents/made>.
- Green, Jeff, Ella Ceron, and Alexandre Tanzi.** 2022. "NYC Aims to Close Stubborn Gender Pay Gap with Salary Disclosure." *Bloomberg*, February 3. <https://www.bloomberg.com/news/articles/2022-02-03/nyc-aims-to-close-stubborn-gender-pay-gap-with-salary-disclosure#xj4y7vzkg>.
- Gulyas, Andreas, Sebastian Seitz, and Sourav Sinha.** 2023. "Does Pay Transparency Affect the Gender Wage Gap? Evidence from Austria." *American Economic Journal: Economic Policy* 15 (2): 236–55.
- Hall, Robert E., and Alan B. Krueger.** 2012. "Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-the-Job Search." *American Economic Journal: Macroeconomics* 4 (4): 56–67.
- Harris, Milton, and Bengt Holmstrom.** 1982. "A Theory of Wage Dynamics." *Review of Economic Studies* 49 (3): 315–33.
- Hauser, Oliver P., and Michael I. Norton.** 2017. "(Mis)perceptions of Inequality." *Current Opinion in*

- Psychology* 18: 21–25.
- Holmström, Bengt.** 1999. “Managerial Incentive Problems: A Dynamic Perspective.” *Review of Economic Studies* 66 (1): 169–82.
- HR Dive.** 2022a. “Salary History Bans: A Running List of States and Localities That Have Outlawed Pay History Questions.” <https://www.hrdiver.com/news/salary-history-ban-states-list/516662> (accessed on November 7, 2023).
- HR Dive.** 2022b. “A Running List of States and Localities That Require Employers to Disclose Pay or Pay Ranges.” <https://www.hrdiver.com/news/pay-transparency-law-tracker-states-that-require-employers-to-post-pay-range-or-wage-range/622542> (accessed on November 7, 2023).
- Hvidberg, Kristoffer B., Claus Kreiner, and Stefanie Stantcheva.** 2020. “Social Positions and Fairness Views on Inequality.” NBER Working Paper 28099.
- Illinois State Comptroller.** 2021. “Salary Database.” Illinois State Comptroller. <https://illinoiscomptroller.gov/financial-reports-data/data-sets-portals/salary-database?IssueYear=2021> (accessed on November 7, 2023).
- International Labour Organization.** 2022. *Pay Transparency Legislation: Implications for Employers’ and Workers’ Organizations*. Geneva, Switzerland: International Labour Organization.
- Jäger, Simon, Christopher Roth, Nina Roussille, and Benjamin Schoefer.** 2021. “Worker Beliefs about Outside Options.” NBER Working Paper 29623.
- Jensen, Robert.** 2010. “The (Perceived) Returns to Education and the Demand for Schooling.” *Quarterly Journal of Economics* 125 (2): 515–48.
- Johanson, Mark.** 2021. “Why Companies Don’t Post Salaries in Job Adverts.” *BBC*, September 22. <https://www.bbc.com/worklife/article/20210921-why-companies-dont-post-salaries-in-job-adverts>.
- Kiatpongsan, Sorapop, and Michael I. Norton.** 2014. “How Much (More) Should CEOs Make? A Universal Desire for More Equal Pay.” *Perspectives on Psychological Science* 9 (6): 587–93.
- Kluegel, James R., and Eliot R. Smith.** 1986. *Beliefs about Inequality: Americans’ Views of What Is and What Ought to Be*. New Brunswick: Aldine Transactions.
- Kreps, David M., and Robert Wilson.** 1982. “Reputation and Imperfect Information.” *Journal of Economic Theory* 27 (2): 253–79.
- Krishna, Vijay.** 2009. *Auction Theory*. Burlington, MA: Academic Press.
- Krueger, Alan B., and Andreas I. Mueller.** 2016. “A Contribution to the Empirics of Reservation Wages.” *American Economic Journal: Economic Policy* 8 (1): 142–79.
- Kuhn, Peter, and Wulong Gu.** 1998. “A Theory of Holdouts in Wage Bargaining.” *American Economic Review* 88 (3): 428–49.
- Kuhn, Peter, and Wulong Gu.** 1999. “Learning in Sequential Wage Negotiations: Theory and Evidence.” *Journal of Labor Economics* 17 (1): 109–40.
- Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva.** 2015. “How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments.” *American Economic Review* 105 (4): 1478–1508.
- Lawler, Edward E.** 1965. “Managers’ Perceptions of Their Subordinates’ Pay and of Their Superiors’ Pay.” *Personnel Psychology* 18 (4): 413–22.
- Lazear, Edward.** 2018. “Compensation and Incentives in the Workplace.” *Journal of Economic Perspectives* 32 (3): 195–214.
- Lazear, Edward P., and Sherwin Rosen.** 1981. “Rank-Order Tournaments as Optimum Labor Contracts.” *Journal of Political Economy* 89 (5): 841–64.
- Littman, Margaret.** 2001. “The Silent Treatment: When it comes to salaries, what women don’t know can hurt them. What they do know can hurt your business.” *Working Woman* 26 (7): 38–43.
- Luttmer, Erzo F. P.** 2005. “Neighbors and Negatives: Relative Earnings and Well-Being.” *Quarterly Journal of Economics* 120 (3): 963–1002.
- Manning, Alan.** 2003. *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton: Princeton University Press. <https://doi.org/10.1515/9781400850679>.
- Mas, Alexandre.** 2016. “Does Disclosure Affect CEO Pay Setting? Evidence from the Passage of the 1934 Securities and Exchange Act.” Princeton University Industrial Relations Section Working Paper 632.
- Mas, Alexandre.** 2017. “Does Transparency Lead to Pay Compression?” *Journal of Political Economy* 125 (5): 1683–1721.
- Matthews, Dylan.** 2019. “In Norway, You Can Look Up Your Neighbor’s Income on the Internet. That’s a Great Idea.” *Vox*, April 8. <https://www.vox.com/2016/5/19/11705746/tax-records-public-norway>.

- Midorikawa, Yoshie.** 2022. “New Rules Require Japanese Companies to Disclose Their Gender Wage Gap.” Society for Human Resource Management, August 4. <https://www.shrm.org/resourcesandtools/hr-topics/global-hr/pages/japan-gender-wage-gap.aspx>.
- Milgrom, Paul R.** 1981. “Good News and Bad News: Representation Theorems and Applications.” *Bell Journal of Economics* 12 (2): 380–91.
- Milgrom, Paul, and Robert Weber.** 1982. “A Theory of Auctions and Competitive Bidding.” *Econometrica* 50 (5): 1089–1122.
- Minami, Takero.** 2018. “Japan to Require Listed Companies to Disclose Executive Pay Rules.” *Nikkei Asia*, December 5. <https://asia.nikkei.com/Business/Companies/Japan-to-require-listed-companies-to-disclose-executive-pay-rules>.
- Minnesota Management and Budget.** 2011. “Payroll Data.” Transparency MN. <https://mn.gov/mmb/transparency-mn/payrolldata.jsp> (accessed on November 7, 2023).
- Mortensen, Dale T.** 1982. “The Matching Process as a Noncooperative Bargaining Game.” In *The Economics of Information and Uncertainty*, edited by John J. McCall, 233–58. Chicago: University of Chicago Press.
- Mortensen, Dale T.** 2005. *Wage Dispersion: Why Are Similar Workers Paid Differently?* Cambridge, MA: MIT Press.
- Mueller, Andreas I., Johannes Spinnewijn, and Giorgio Topa.** 2021. “Job Seekers’ Perceptions and Employment Prospects: Heterogeneity, Duration Dependence, and Bias.” *American Economic Review* 111 (1): 324–63.
- Nadworny, Katie.** 2022. “Israel: Equal Pay Reporting Obligations.” Society for Human Resource Management, May 12. <https://www.shrm.org/resourcesandtools/hr-topics/global-hr/pages/israel-equal-pay-reporting-obligations.aspx>.
- National Women’s Law Center (NLWC).** 2020. “Promoting Pay Transparency to Fight the Gender Wage Gap.” <https://nwlc.org/wp-content/uploads/2018/06/International-Pay-Transparency-Models-v2.pdf>.
- New Jersey Department of Labor (NJDOLE).** 2018. New Jersey Equal Pay Act. <https://www.nj.gov/oag/dcr/downloads/Equal-Pay-Act.pdf>.
- Norton, Michael I., and Dan Ariely.** 2011. “Building a Better America—One Wealth Quintile at a Time.” *Perspectives on Psychological Science* 6 (1): 9–12.
- Obama, Barack.** 2014. “Executive Order—Non-retaliation for Disclosure of Compensation Information.” Executive Order 13665. <https://obamawhitehouse.archives.gov/the-press-office/2014/04/08/executive-order-non-retaliation-disclosure-compensation-information>.
- Obloj, Tomasz, and Todd Zenger.** 2022. “The Influence of Pay Transparency on (Gender) Inequity, Inequality, and the Performance Basis of Pay.” *Nature Human Behavior* 6 (5): 646–55.
- Ockenfels, Axel, Dirk Sliwka, and Peter Werner.** 2015. “Bonus Payments and Reference Point Violations.” *Management Science* 61 (7): 1496–1513.
- O’Donoghue, Julie.** 2023. “Louisiana Will Make State Government Salaries Available to the Public.” *Louisiana Illuminator*, February 21. <https://lailluminator.com/briefs/louisiana-will-make-state-government-salaries-available-to-the-public/>.
- Percivalle, Uberto.** 2021. “Italy: Gender-Equality Reporting and Certification Bill Passed.” Society for Human Resource Management, November 16. <https://www.shrm.org/resourcesandtools/hr-topics/global-hr/pages/italy-gender-equality-reporting-and-certification.aspx>.
- Perez-Truglia, Ricardo.** 2020. “The Effects of Income Transparency on Well-Being: Evidence from a Natural Experiment.” *American Economic Review* 110 (4): 1019–54.
- Perez-Truglia, Ricardo, and Ugo Troiano.** 2018. “Shaming Tax Delinquents.” *Journal of Public Economics* 167: 120–37.
- Phillips, Macon.** 2009. “Now Comes Lilly Ledbetter.” <https://obamawhitehouse.archives.gov/blog/2009/01/25/now-comes-lilly-ledbetter>.
- Pissarides, Christopher A.** 1985. “Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages.” *American Economic Review* 75 (4): 676–90.
- Postel-Vinay, Fabien, and Jean-Marc Robin.** 2002. “The Distribution of Earnings in an Equilibrium Search Model with State-Dependent Offers and Counteroffers.” *International Economic Review* 43 (4): 989–1016.
- Reese, Phillip.** 2019. “See How Far Union Membership Has Declined in California.” *Sacramento Bee*, January 29. <https://www.sacbee.com/news/california/article225087150.html>.
- Robinson, Joan.** 1933. *The Economics of Imperfect Competition*. London, UK: Macmillan.

- Rosen, Sherwin.** 1986. "Prizes and Incentives in Elimination Tournaments." *American Economic Review* 76 (4): 701–15.
- Roussille, Nina.** 2021. "The Central Role of the Ask Gap in Gender Pay Inequality." Unpublished.
- Shepherd, Leah.** 2022. "Ireland Requires Employers to Report Pay Disparities by Gender." Society for Human Resource Management, May 9. <https://www.shrm.org/resourcesandtools/hr-topics/global-hr/pages/ireland-pay-disparities-by-gender.aspx>.
- Skoda, Samuel.** 2022. "Directing Job Search in Practice: Mandating Pay Information in Job Ads." Unpublished.
- Spinnewijn, Johannes.** 2015. "Unemployed but Optimistic: Optimal Insurance Design with Biased Beliefs." *Journal of the European Economic Association* 13 (1): 130–67.
- State of New Hampshire.** 2009. "State Employee Pay Search." Transparent NH. <https://business.nh.gov/paytransparency/PayTransparencyResults.aspx> (accessed on November 7, 2023).
- State of New Jersey Transparency Center.** 2010. "Public Payroll." State of New Jersey Department of the Treasury. <https://www.nj.gov/transparency/payroll> (accessed on November 7, 2023).
- State of Oregon.** 2016. "State Salaries" Oregon Transparency. <https://www.oregon.gov/transparency/pages/state-salaries.aspx> (accessed on November 7, 2023).
- State of Vermont.** 2017. "State of Vermont Employee Salaries." Vermont Data Portal. <https://data.vermont.gov/Government/State-of-Vermont-Employee-Salaries/jgqy-2smf> (accessed on November 7, 2023).
- Stigler, George J.** 1962. "Information in the Labor Market." *Journal of Political Economy* 70 (5 Part 2): 94–105.
- Sun, Shengwei, Jake Rosenfeld, and Patrick Denice.** 2021. "On the Books, off the Record: Examining the Effectiveness of Pay Secrecy Laws in the U.S." Institute for Women's Policy Research Report C494.
- Toronto Stock Exchange.** 2020. *Guide to Security Based Compensation Arrangements*. Toronto: TSX Inc.
- Trachtman, Richard.** 1999. "The Money Taboo: Its Effects in Everyday Life and in the Practice of Psychotherapy." *Clinical Social Work Journal* 27 (3): 275–88.
- TWSE.** 2021. Securities and Exchange Act Article 14. <https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=G0400001>.
- US Census Bureau.** 2020. "Post-secondary Employment Outcomes (PSEO)." [https://lehd.ces.census.gov/data/pseo\\_experimental.html](https://lehd.ces.census.gov/data/pseo_experimental.html) (accessed on November 7, 2023).
- US Securities and Exchange Commission.** 2015. "SEC Adopts Rule for Pay Ratio Disclosure." <https://www.sec.gov/news/press-release/2015-160>.
- US Securities and Exchange Commission.** 2022. "SEC Adopts Pay versus Performance Disclosure Rules." <https://www.sec.gov/news/press-release/2022-149>.
- Valdez, José Antonio, and Lucianna Polar.** 2018. "Peru: New Legislative Protection for Equal Pay." Society for Human Resource Management, February 16. <https://www.shrm.org/resourcesandtools/legal-and-compliance/employment-law/pages/global-peru-equal-pay-legislation.aspx>.
- Williams, Steven R.** 1987. "Efficient Performance in Two Agent Bargaining." *Journal of Economic Theory* 41 (1): 154–72.
- Wiswall, Matthew, and Basit Zafar.** 2014. "Determinants of College Major Choice: Identification Using an Information Experiment." *Review of Economic Studies* 82 (2): 791–824.
- Women's Bureau.** 2022. "Equal Pay and Pay Transparency Protections." US Department of Labor. <https://www.dol.gov/agencies/wb/equal-pay-protections> (accessed on November 7, 2023).



# Immigration and Crime: An International Perspective

Olivier Marie and Paolo Pinotti

In the late 1920s, President Herbert Hoover appointed a National Commission on Law Observance and Enforcement to address rising concerns about crime. The Commission dedicated one of its final reports to the issue of “Crime and the Foreign Born,” which begins by noting that “[t]he theory that immigration is responsible for crime . . . is almost as old as the colonies planted by Englishmen on the New England coast” (National Commission on Law Observance and Enforcement 1931, p. 23).

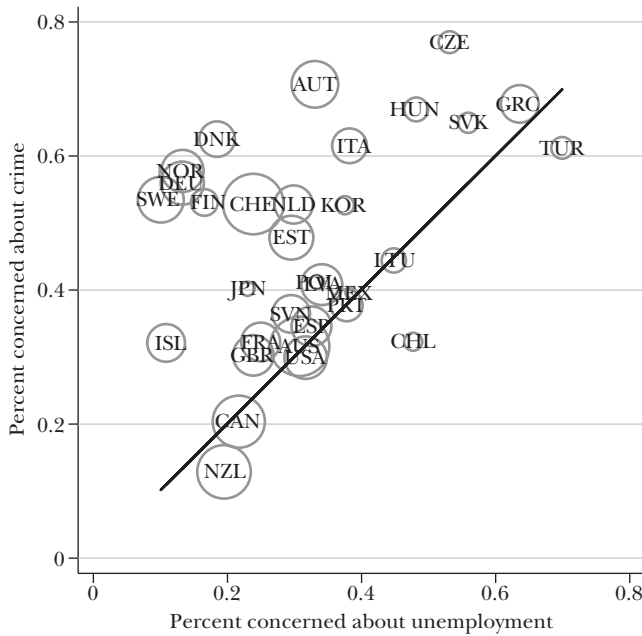
Almost a century later, concerns about crime remain one of the most frequently expressed reasons for public opposition to immigration in countries around the world. Figure 1 plots the share of respondents in OECD countries worried that “immigrants increase crime” against the share of respondents worried that “immigrants take jobs away from natives,” which is another prominent public opinion concern (for example, Mayda 2006; Haaland and Roth 2020). Most countries lie

■ *Olivier Marie is Professor of Labour Economics, Erasmus School of Economics, Rotterdam, Netherlands, and Professor of Applied Labour Economics, Research Centre for Education and the Labour Market at Maastricht University, Maastricht, Netherlands. He is also a Research Fellow, Tinbergen Institute, Amsterdam and Rotterdam, Netherlands; Network Affiliate, Ifo Institute and Center for Economic Studies, Munich, Germany; and Research Fellow, Institute for the Study of Labor (IZA), Bonn, Germany. Paolo Pinotti is Professor of Economics and Endowed Chair in Economic Analysis of Crime in the Department of Social and Political Sciences at Bocconi University, where he also Directs of the CLEAN Unit on Economic Analysis of Crime of the BAFFI Research Center. Both authors are Research Affiliates, Centre for Economic Policy Research (CEPR), London, United Kingdom. Their email addresses are marie@ese.eur.nl and paolo.pinotti@unibocconi.it.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.1.181>.

Figure 1

**Natives' Concerns in OECD Countries about the Effect of Immigration on Crime and Unemployment**



Source: Authors' compilation from the 2017–2020 waves of the World Values Survey and European Values Survey (EVS 2023; Haerpfer et al. 2022).

Notes: Concerns about crime are measured by the share of respondents that agree with the statement “Immigrants increase crime.” Concerns about unemployment are measured by the share of respondents that agree with the statement “Immigrants increase unemployment.” We included all countries in the European Values Survey and all other OECD countries in the World Values Survey: Australia (AUS), Austria (AUT), Chile (CHL), Czechia (CZE), Denmark (DNK), Estonia (EST), Finland (FIN), France (FRA), Germany (DEU), Great Britain (GBR), Greece (GRC), Hungary (HUN), Iceland (ISL), Italy (ITA), Japan (JPN), Lithuania (LTU), Mexico (MEX), Netherlands (NLD), New Zealand (NZL), Norway (NOR), Poland (POL), Portugal (PRT), Slovakia (SVK), Slovenia (SVN), South Korea (KOR), Spain (ESP), Sweden (SWE), Switzerland (CHE), Türkiye (TUR), United States (USA). The size of the markers is proportional to the share of immigrants over total population. The 45-degree line is also shown in the graph.

well above the 45-degree line, meaning that more people worry about the crime effects of immigration than about its labor market effects. This pattern holds especially for countries with a higher share of migrants over total population, indicated by larger circles.

Evidence on the evolution of beliefs concerning immigrants and crime over time can be documented using survey evidence from the United States. For the past 30 years, Gallup has regularly surveyed a representative sample of Americans, asking



them to consider various topics and “say whether immigrants to the United States are making the situation in the country better or worse, or not having much effect.”<sup>1</sup> When the topic is “the crime situation,” half or more of the Americans surveyed have answered “worse” in almost every wave. If these views have lately somewhat improved, with 42 percent of respondents choosing “worse” in 2019, this answer was still 18 points higher than the proportion answering that immigrants worsen their job opportunities in the same year.

Perhaps surprisingly, economists have only recently begun to investigate the links between immigration and crime in a systematic way. Advances in methodology and data quality during the past few decades have made it possible to go beyond simple correlations and to assess the causal impact of immigration on crime. This evolution has had much in common with the study of the labor market effects of migration depicted in this journal by Peri (2016).

In this paper, we first describe recent international trends in immigration and crime, exploring why migrants may at first appear much more criminally active in certain countries. We then discuss the theoretical framework and methodological tools that economists have used in thinking about the relationship between immigration and crime. We assess what these approaches have produced in various contexts as to the causal impact of immigration levels on crime rates. We review the evidence on this point so far, which overwhelmingly suggests that immigrants do not increase crime levels in the communities where they settle, and confirm this overall null-effect conclusion using newly collected international data. Finally, we consider the evidence on the links between access to legal work and the crime propensity of different kinds of immigrants, including refugees and those with irregular legal status. The relatively few papers that have probed this issue all conclude that legal status and work permits strongly decrease the probability that immigrants will become involved in crime.

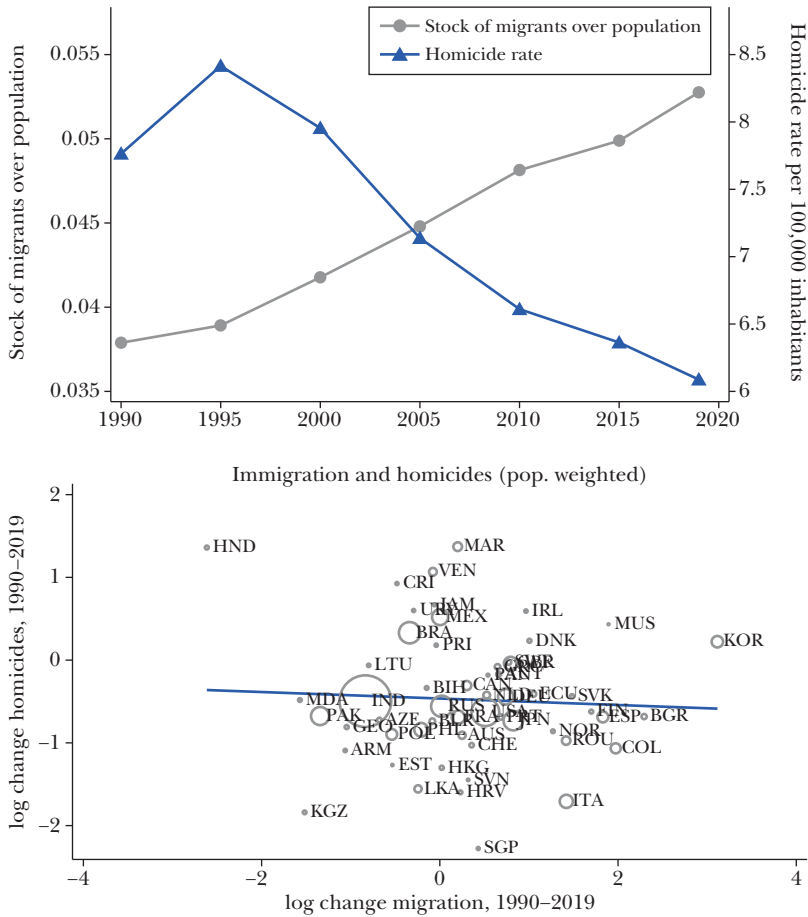
## **International Patterns of Immigration and Crime**

### **Correlating Immigration and Crime across Countries**

The top panel of Figure 2 plots the change in the number of immigrant arrivals and of homicides recorded per population in 55 countries over the period 1990–2019. We focus on homicide data because it is comparable across countries, owing to the limited underreporting (if any) for this type of crime compared to other crime categories, such as property or even other violent crime. This graph clearly shows a negative relationship between immigration and crime over the past

<sup>1</sup>Individuals surveyed by Gallup in 2001, 2002, 2004, 2007, 2017, and 2019, were asked the following question: “For each of the following areas, please say whether immigrants to the United States are making the situation in the country better or worse, or not having much effect.” Respondents could answer either better, worse, not much effect, or no opinion. Among the topics to consider were “the crime situation” and about “job opportunities for you/your family” (<https://news.gallup.com/poll/1660/immigration.aspx>).

**Figure 2**  
**Immigration and Homicides in 55 Countries, 1990–2019**



*Source:* Authors' own compilation from combining the United Nations' Global Migration Data (United Nations 2023a) and Global Study on Homicide (United Nations 2023b).

*Notes:* The graph in the top panel plots the average homicide rate and the average share of foreigners over total residents across 55 countries: Armenia (ARM), Australia (AUS), Austria (AUT), Azerbaijan (AZE), Bulgaria (BGR), Bosnia and Herzegovina (BIH), Belarus (BLR), Brazil (BRA), Canada (CAN), Switzerland (CHE), Colombia (COL), Costa Rica (CRI), Germany (DEU), Denmark (DNK), Ecuador (ECU), Spain (ESP), Estonia (EST), Finland (FIN), France (FRA), United Kingdom (GBR), Georgia (GEO), Greece (GRC), Hong Kong (HKG), Honduras (HND), Croatia (HRV), India (IND), Ireland (IRL), Italy (ITA), Jamaica (JAM), Japan (JPN), Kyrgyzstan (KGZ), South Korea (KOR), Sri Lanka (LKA), Lithuania (LTU), Morocco (MAR), Moldova (MDA), Mexico (MEX), Mauritius (MUS), Netherlands (NLD), Norway (NOR), Pakistan (PAK), Panama (PAN), Philippines (PHL), Poland (POL), Puerto Rico (PRI), Portugal (PRT), Romania (ROU), Russia (RUS), Singapore (SGP), Slovakia (SVK), Slovenia (SVN), Sweden (SWE), Uruguay (URY), United States of America (USA), and Venezuela (VEN). The top panel covers the period 1990–2019 (in five-year intervals). The bottom panel shows the cross-country relationship between the (log) change of the two variables over the same period.

30 years, as the average homicide rate across countries dropped by one-third—from a peak of 8.5 per 100,000 inhabitants in 1995 to just above 6 in 2019—while the share of foreign residents increased by two thirds.

Of course, these average international trends may mask important differences in the relationship between changes in immigration and crime across countries. For this reason, in the bottom panel of Figure 2 we plot the (log) change in homicide rates between 1990 and 2019 (on the y-axis) against the (log) change in immigration over the same period (on the x-axis) across the 55 countries for which we have these data. The size of the circles indicates the relative population size of each country. The regression line, also shown in the graph, is not significantly different from zero.

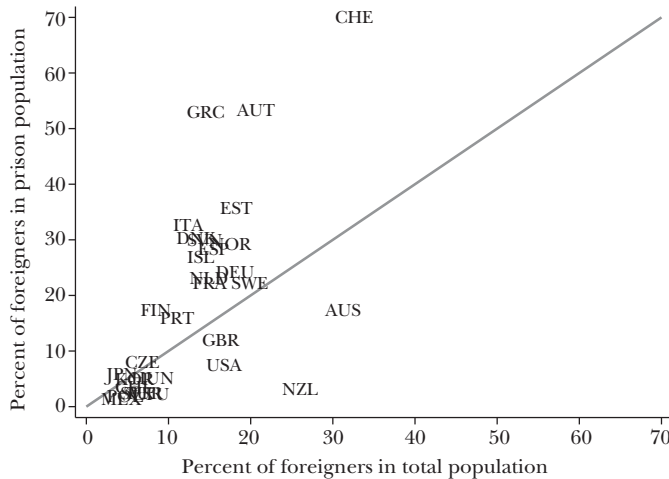
From this exercise, there appears to be no correlation between immigration and crime, at least at the cross-national level. Of course, these correlations should not be taken as causal evidence of an immigration-to-crime relationship, something for which we need to use the methodological tools we explain in detail later; nevertheless, they suggest that broad trends in immigration and crime are not aligned with the high levels of public worry documented earlier. Is this the result of important misperceptions about the level of participation of foreign nationals in criminal activity—something quite common when it comes to numbers linked to immigration and which can be improved with information (Grigorieff, Roth, and Ubfal 2020; Alesina, Miano, and Stantcheva 2022)—or is there other evidence that may tend to support this perceived overrepresentation?

### **The Overrepresentation of Migrants in Crime**

Comparing the prevalence of crime across native-born and migrant populations across countries is not easy, but one possible approach uses prison population numbers. One reason why people might believe that immigrants worsen the crime situation may stem from the fact that, in most developed countries, the share of foreign-born prisoners surpasses—often by a large extent—the share of foreign-born residents. Notable exceptions in this respect are the United States and a few other Anglo-Saxon countries: Abramitzky et al. (2023) provide evidence of the underrepresentation of immigrants in US prisons over several decades.

The frequent overrepresentation of incarcerated immigrants relative to overall population, depicted in Figure 3, might be due to two measurement issues quite specific to immigration-crime statistics. One is that irregular aliens (that is, those not officially registered by the appropriate process in that nation) would be counted in the share of immigrants in the prison population, but not in the share of immigrants in the resident population. The other is the potential harsher treatment of foreigners by the police and the judiciary system due to discrimination or unequal access to legal services and noncustodial measures, such as bail or home-detention. Still, these factors cannot plausibly account for the high incarceration rate of immigrants, which is 2 times higher than that of natives on average across these countries, up to 3 times higher in Austria and Denmark, and 4.5 times higher in Greece.

Figure 3

**Overrepresentation of Foreigners in Prison Populations**

*Source:* Authors' own compilation combining data on the proportion of foreign prisoners from ICPR (2023) and data on proportion of foreign residents in the same country in the same year from the OECD (2023).

*Notes:* The countries listed are Australia (AUS), Austria (AUT), Chile (CHL), Czechia (CZE), Denmark (DNK), Estonia (EST), Finland (FIN), France (FRA), Germany (DEU), Great Britain (GBR), Greece (GRC), Hungary (HUN), Iceland (ISL), Italy (ITA), Japan (JPN), Lithuania (LTU), Mexico (MEX), Netherlands (NLD), New Zealand (NZL), Norway (NOR), Poland (POL), Portugal (PRT), Slovakia (SVK), Slovenia (SVN), South Korea (KOR), Spain (ESP), Sweden (SWE), Switzerland (CHE), Türkiye (TUR), and the United States (USA). The data refer to the latest year available, ranging between 2014 and 2020. The 45-degree line is also shown in the graph.

A more general reason as to why immigrants appear to disproportionately contribute to prison populations in most countries may come from important differences in the composition of the foreign and native populations in terms of basic characteristics such as gender, age, and education. Foreigners are overrepresented in the young male population of most countries, and we know that men are much more criminally active than women and that offending behavior has long been observed to peak during an individual's late teens and early 20s (for example, Quetelet 2003; Hirschi and Gottfredson 1983). Moreover, in many destination countries, immigrants are on average less-educated than natives. The OECD classifies foreign-born and native-born in each country as either having low or high education levels in its 2018 Indicators of Immigrant Integration report (OECD 2018). In the EU28 group of countries, 22.5 percent of natives are classified as low-educated, compared to 33.9 percent of foreign-born; in the United States, only 7.5 percent of natives are low-educated, compared to 23.5 percent of foreign-born. Because there is now very strong causal evidence of a positive effect of education in preventing criminal participation (Lochner and Moretti 2004; Machin, Marie, and Vujčić 2011), this is another factor that may contribute to the overrepresentation

of immigrants among offenders. The impact this education gap has on immigrant criminal propensity may be context-specific and depend on the importance of “skill downgrading,” which immigrants limits to jobs below their skill level because of language or legal barriers (for a review in this journal, see Dustmann, Schönberg, and Stuhler 2016), for the earnings potential of immigrants in different countries.

While comparing crime rates among immigrants and natives—either conditional on other individual characteristics or not—is informative, it may not be the most relevant comparison from a policy perspective. It is probably more pertinent to ask whether higher migration causes an increase in crime rates, which is what citizens, voters, and decision-makers in destination countries care about. Of course, immigrants’ involvement in crime and the impact of immigration on aggregate crime rates are intimately related. To the extent that immigrants display a higher probability than natives of committing crimes, one may expect that higher immigration inflows would lead to increases in measurable criminal activity in destination countries. In practice, this does not need to be the case if, for example, the share of migrants in the population is too small to influence crime rates or if they directly substitute for and displace natives’ criminal activity. Evaluating the relationship between immigration and crime is thus more complex than the simple interpretation of correlation and raw statistics, requiring some theoretical framing as well as state-of-the-art econometric tools.

## **Theoretical Framework and Empirical Challenges**

The workhorse Becker (1968) model of criminal participation decision-making—based on an individual cost-benefit analysis of whether to commit crimes—delivers some straightforward predictions regarding immigrants’ involvement in crime relative to the native-born population. First, crime propensity is expected to be higher among immigrants relative to the local population if immigrants are relatively negatively selected in terms of certain characteristics linked to legal labor market earnings potential, such as education and skill level. Second, immigrants are more likely to commit crimes if they face worse wage or employment prospects than their demographic characteristics would predict, due to labor market discrimination, skill downgrading, or illegal status.

These factors suggest that if immigrants arriving in an area are disproportionately young, unskilled males facing some form of labor market barrier to entry, then one could reasonably expect that crime rates would increase. But although this belief is probably one reason behind the strong link in public opinion between immigration and crime, it is too simplistic for two main reasons. First, the potential costs associated with criminal activity for foreign residents may be higher than for natives—higher arrest probability, harsher sentences, and the risk of deportation—leading fewer immigrants to choose to participate. Second, as we move from the individual crime choice to local crime rates, the latter may depend on general equilibrium effects linked to congestion effects in labor markets and welfare (for

example, how migrant inflows complement some native-born workers but substitute for others), interactions in crime, social multipliers, and so on. Given these issues, theoretical conclusions about the relationship between immigration and crime have an ambiguous sign, and the relationship is ultimately an empirical issue.

However, a combination of selection and measurement issues make it difficult to estimate a causal effect of immigration on local crime rates. The selection problem is common to research into economic impacts of migration that must account for foreign arrivals moving disproportionately into booming (for jobs) or depressed (for housing) locations. For example, if migrants choose economically expanding areas with comparatively low crime rates, then a simple ordinary least squares correlation across localities at a point in time will tend to overstate how immigration is related to job growth and understate how it is related to crime. A more sophisticated approach might look at data about immigration and localities over time, including control variables for local economic activity as well as area and time fixed effects. While a more careful approach along these lines should greatly reduce the potential bias, compared to a basic ordinary least squares approach, it would still not necessarily yield causal estimates, for two main reasons. First, unobservable time-varying, area-specific factors that affect both immigration and crime may remain, such as changes in local policies (for example, measures that are simultaneously anti-migration and pro-police) or housing prices. Second, area crime dynamics may themselves influence migrant location choice, creating other reverse causality issues.

To address these issues, economists have often used the shift-share instrumental variable approach. This approach rose to prominence after it was used by Bartik (1991) to investigate the effects of local economic growth policies and was adapted by Altonji and Card (1991) in the migration context. A decade later, it was further developed as a formal tool to investigate the causal labor market impact of immigration by Card (2001) and has since become the norm in this literature. This strategy is based on an assumption that the location decision of new migrants with a specific nationality (that is, *shifts* in number of new arrivals) is causally affected by the historical location of previously-arrived migrants from the same nationality across areas (that is, the *share* of immigrants from a given nationality in each country, region, district, and so on). If this assumption holds true, it can be used as the first stage in a two-stage least squares approach. In the first stage, the historical share of immigration across localities is used to predict current immigration. In the second, the predicted (and thus arguably exogenous) values of immigration in a given area from the first-stage regression are used to predict crime in that area.

The validity of the shift-share instrumental variable strategy relies on initial migrant distribution from specific countries not being correlated with persistent area-specific factors that affect local crime rates. This crucial “exclusion restriction” is not directly testable but can be made more convincing by looking at the location of the previously-arrived stock of immigrants several years before the period of analysis, and, preferably, prior to a relatively important increase in migrant flows. A simple validity exercise is to check for (no) correlation between initial shares of immigrants from a certain area in a given origin country and changes in area-specific factors

that can influence crime, or crime rates themselves, in the pre-analysis period. Additional robustness checks have recently been proposed that can further reinforce the validity of the shift-share instrumental approach for calculating causal estimates of complex and sometimes elusive relationships, including that between immigration and crime (Jaeger, Ruist, and Stuhler 2018; Adão, Kolesár, and Morales 2019; Goldsmith-Pinkham, Sorkin, and Swift 2020; Borusyak, Hull, and Jaravel 2022).

The measurement problem is more pronounced in the study of the migration–crime relationship than in most other dimensions of the economics of migration. The problem is that the official numbers on *both* migration and crime may be an unreliable reflection of reality, and simultaneously stem from policy decisions that affect both measures. Countries, states, or municipalities with relatively stronger institutions may consistently produce more reliable crime and migration metrics. The extent of local policy focus on law and order may affect both variables simultaneously. Such problems can mostly be addressed empirically or by using secondary data. For example, if the downward (or upward) bias in crime or migration statistics is relatively *constant*—say only half of crimes are reported, and half the migrants are legally registered—then comparing numbers over time should still reflect real changes. Also, events in which undocumented migrants are given amnesty can be used to check official numbers against the actual migrant population and, in particular, to confirm their geographical distribution. Rates at which crimes are “cleared” with an arrest referral to court, a standard measure of police productivity, can be used as a proxy for changes in an area’s law and order focus (as suggested by Ehrlich 1996). Most studies that have convincingly investigated the empirical relationship between immigration and crime have attempted to address these measurement issues.

## Empirical Estimates of How Immigration Affects Crime Rates

### Current National-Level Evidence

One of the first studies to estimate the causal effect of immigration on crime, conducted by Butcher and Piehl (1998), used US data across 43 metropolitan areas during the period 1979–1990. They first show, using a standard ordinary least squares regression, that immigration is positively correlated with local crime rates. However, when they use the share of foreign-born in 1979 as an instrument for the inflows during the period 1980–1990, the estimated causal effect on the local crime rate is not significantly different from zero.<sup>2</sup> Spenkuch (2014) obtains similar results using US county-level data for the period 1980–2000. Both ordinary least squares and shift-share instrumental variable estimates of the crime elasticity are very small in magnitude (0.11 and 0.01 for property and violent crimes, respectively) and the causal shift-share estimates are not statistically distinguishable from zero.

<sup>2</sup>Butcher and Piehl (1998) only exploit variation from the “share” component of the “shift-share” instrumental variable, following Altonji and Card (1991).

These results for the United States are confirmed by studies using the same empirical strategy to estimate the impact of immigration on crime in other countries. Looking across Italian provinces for the period 1990–2003, Bianchi, Buonanno, and Pinotti (2012) show that immigration is positively correlated with property and violent crimes, but the causal effect is not significantly different from zero. To counter potential measurement issues, they account for the presence of undocumented immigrants by exploiting data on the numerous amnesties of irregular immigrants implemented during this period in Italy, though they cannot disentangle the impact of regular and irregular migration due to the extremely high correlation between the two across provinces. At the same time, regular and irregular migrants have very different opportunities in the labor market, which in turn would affect their opportunity cost of committing crime.

Immigrants are a heterogeneous group: one key difference is between economic migrants and asylum-seekers. Looking at UK data, Bell, Fasani, and Machin (2013) estimate the impact of two different groups of immigrants: citizens of eight Eastern European countries admitted to the European Union in 2004 and asylum seekers. Migrants within the European Union had rights to live and work in the United Kingdom, which should lower the propensity to engage in criminal activities. Consistent with this hypothesis, their impact on the overall crime rate—estimated using a shift-share instrumental variable—is not significantly different from zero, and the estimated impact on property crimes is actually negative. For asylum-seekers, who experienced worse earning opportunities and were in many cases denied permanent residence status, Bell, Fasani, and Machin estimate the crime effect by exploiting the quasi-random dispersal policy implemented by the UK government. We discuss these results in the next main section with other papers that also use quasi-random allocation of migrants for causal identification.

Overall, the evidence from shift-share instrumental variable estimates in the United States and in European countries suggests no significant effect of immigration on property or violent crimes. Ajzenman, Dominguez, and Undurraga (2023) confirm this result in another context—immigration to Chile—using victimization rates, instead of police-recorded offenses, as their measure of crime.

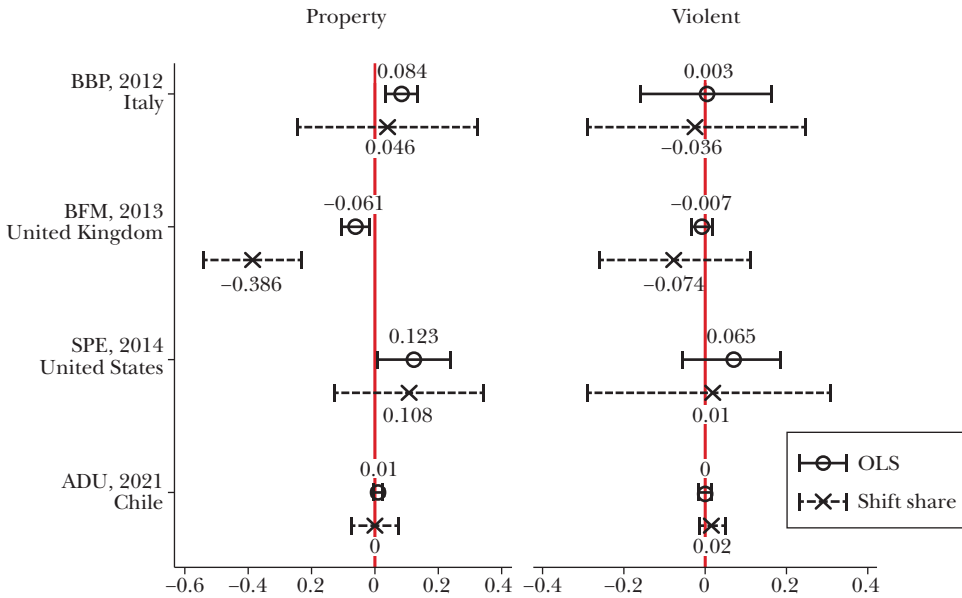
Figure 4 summarizes the available estimates of the elasticity of immigration from the papers discussed in this section that provide both ordinary least squares and shift-share instrumental variable estimates of how immigration affects both property and violent crimes. The figure shows coefficients and the 95 percent confidence interval, with the vertical red line indicating no effect. In all contexts, the causal estimates produced by the shift-share methods are not significantly different from zero (and negative in one case). Importantly, the figure also shows that the instrumental variable estimates are invariably more negative than ordinary least squares estimates—rather than just being less precisely estimated.

Overall, there seem to be no discernable causal effect of increased migration inflows in an area on the number of property or violent offences it experiences. Next, we look at a new level of geographical analysis: international evidence across European countries and regions.



Figure 4

Summary of Estimates of the Impact of Immigration on Crime



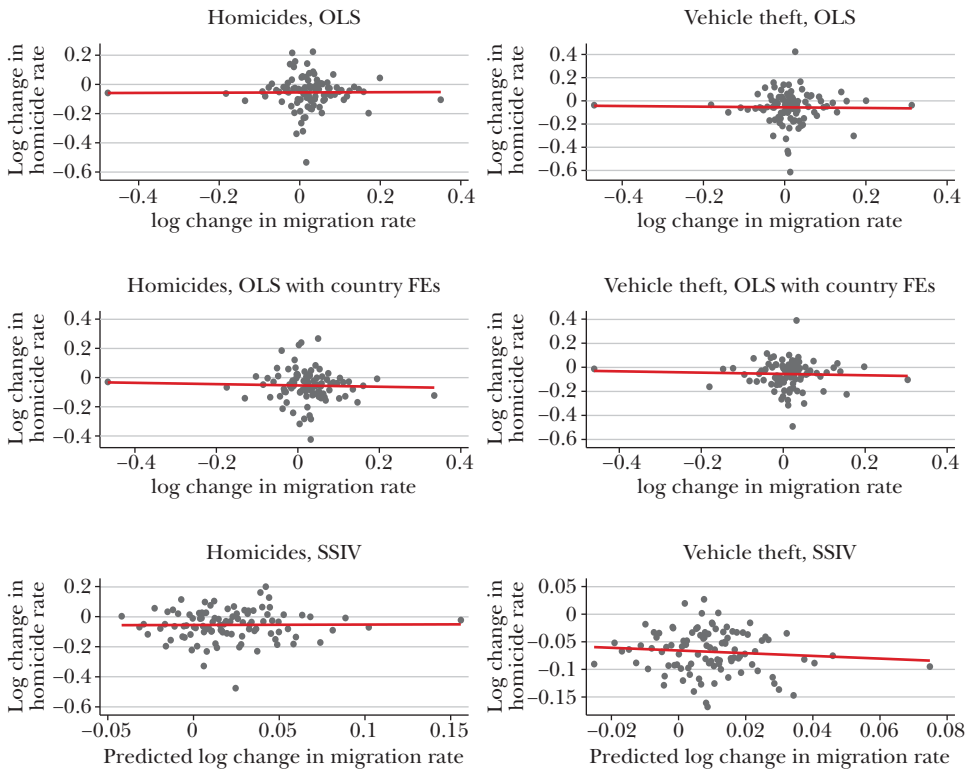
Source: The full references of the papers from which these estimates are taken from are: BBP 2012: Bianchi, Buonanno, and Pinotti (2012); BFM 2013: Bell, Fasani, and Machin (2013); SPE 2014: Spenkuch (2014); ADU 2021: Ajzenman, Dominguez, and Undurraga (2023).

Notes: The two panels of this figure summarize the elasticity response of property and crime rates with respect to a 1 percent increase in the share of immigrants obtained in papers that use a shift share approach to obtain causal estimates in various contexts. The full lines with the hollow circle represent the ordinary least squares estimates with 95 percent confidence intervals, and the dotted lines with a cross the causal shift-share estimates with their 95 percent confidence intervals.

New International Evidence

In the research reviewed so far, each study focuses on a single country. This approach has some clear advantages, but it could hide patterns emerging across countries. For this reason, we complement these previous findings with novel evidence from comparable data across European countries and regions. Specifically, we assembled data on migration and crime for a yearly panel of 216 regions (level 2 of the Nomenclature of Territorial Units for Statistics, or NUTS) from 23 European countries over the period 2002–2017, which are publicly available from Eurostat and OECD. These regions divide EU countries into geographical areas with populations typically between 800,000 and 3 million. Figure 5 plots the relationship, across regions, between the log changes in the share of foreign residents over the sample period 2002–2017, on the horizontal axis, and the log changes in homicide rates (left panels) and motor vehicle thefts (right panels) over the same period. We focus

Figure 5

**Immigration and Crime across European Regions**

*Source:* The homicides and vehicle theft data are from Eurostat (2023), the migration and population numbers from the OECD (2023), and initial migrant shares to build the SSIV from Alesina et al (2021). *Note:* The figure plots the log changes in homicide rates (left panel graphs) and motor vehicle thefts (right panel graphs) during the period 2002–2017 against log changes (top two panels) or predicted log changes (bottom panel) in the rate of immigrants over total population during the same period across European regions. The graphs present binned scatters of the mean relationship between the two variables (100 equal bins) and associated fitted regression line for three separate specifications. The top panel is from an ordinary least squares (OLS) regression with year fixed effects and the middle panel from an OLS regression that adds country fixed effects (FEs). The bottom panel results from the shift-share instrumental variable (SSIV) approach that regresses the predicted log change in migration (resulting from the first stage regression) on the log change in the crime rate per crime type. The SSIV regressions include year and country fixed effects. All specifications, OLS and SSIV, are weighted for baseline population size in 2002 and control for log changes in GDP per capita and population. The original data cover 216 European regions at level-2 of the Nomenclature of Territorial Units for Statistics (NUTS2) in 23 countries. We use data for the 73 (85) regions in 11 (10) countries for which the homicide (vehicle theft) numbers are consistently reported over this period.

the analysis on these two types of crime because they are consistently defined across countries and more accurately measured than other types of crimes.

The top two panels show the ordinary least squares (OLS) correlations. The univariate regression of (log changes in) homicide and property crime rates on

immigration is flat (top graphs). Absorbing national-level shocks through the inclusion of country fixed effects, as we do in the middle graphs of Figure 5, enables us to focus on variation across regions within the same country. The relationship between immigration and crime now becomes somewhat negative, though it is far from being statistically significant.

The bottom two panels show the results of a shift-share instrumental variable (SSIV) approach. In these panels, the change in the immigration rate on the horizontal axis represents the *predicted* migration in a region between 2002 and 2017, based on migrant shares in these regions in 2000, and national shifts in migration (in- and outflows) at the national level. The estimates are noisier using this causal approach, but there is still no indication that increases in immigration are systematically associated with increases in crime.<sup>3</sup>

### **Crime and the Access of Migrants to Legitimate Income Opportunities**

The evidence presented in the previous section focuses on immigrants as a single group. However, this approach may conceal significant variation across groups, as foreign migrants represent a very heterogeneous population. In particular, the access that migrants have to legitimate income opportunities may vary for, say, legal economic migrants, legal asylum-seekers, and undocumented migrants. Such differences may be particularly relevant for the purpose of our analysis, as the opportunities for earning legitimate income will affect the opportunity cost of committing crimes.

We mentioned earlier the study of two types of migrants to the United Kingdom by Bell, Fasani, and Machin (2013): citizens of the eight Eastern European countries admitted to the European Union in 2004, and asylum-seekers. Areas in the United Kingdom where the asylum-seekers were quasi-randomly allocated, because of the country's dispersal policy, experienced significantly more property crimes, with a 1 percent increase in refugees among the local population leading to a 1.1 percent increase in police-recorded property crimes. On the other hand, there was no crime increase from the influx of immigrants from new EU member countries. These results are consistent with the fact that new EU citizens had full access to legitimate income opportunities in official labor market—and, thus, a higher opportunity cost of committing crimes—while asylum seekers could not work in the host country.<sup>4</sup> Indeed, asylum-seekers often face significant barriers to employment in the host

<sup>3</sup>The regression results that correspond to the six panels of Figure 5 can be found in Table A1 of the online Appendix. Data and coding for replicating the figure are available as supplementary material on the JEP website.

<sup>4</sup>The results of Bell, Fasani, and Machin (2013) are consistent with those of Piopiunik and Ruhose (2017), who study the impact of ethnic Germans returning to Germany after the collapse of the Soviet Union. They do not detect any effect on property crime for this group, who arrived in the host country with all same rights as native citizens.

countries, which have been shown to have important negative effects on their long-term labor market prospects (Fasani, Frattini, and Minale 2021). Access to and levels of social transfer available may also matter as, for example, Andersen, Dustmann, and Landersø (2019) found that refugee benefit cuts in Denmark increased refugees' involvement in crime.

More recently, some papers have focused on the crucial role of legal status as a determinant of migrants' criminal participation behavior in host countries. To estimate the causal effect of legal status, one would ideally allocate it at random and then compare the probability of committing crime between legalized and nonlegalized immigrants. Although this experimental design is infeasible, many policies approximate this ideal experiment. One example is the accession of Romania and Bulgaria to the European Union on January 1, 2007. As a consequence, all citizens of these countries obtained legal status—including access to the labor market—in all other EU member states. Mastrobuoni and Pinotti (2015) compare the recidivism of Romanian and Bulgarian prison inmates pardoned in Italy with a Collective Clemency Bill on August 2006 (that is, five months before the EU enlargement) with the recidivism of pardoned inmates from countries still awaiting membership to the European Union, before and after the enlargement. They conclude that legal status matters, with recidivism of Romanians and Bulgarians living in Italy declining by over 50 percent relative to the recidivism of other pardoned inmates after the enlargement.

A different policy experiment in Italy involves the online procedure used to award work permits. Prospective employers of immigrant workers must send an electronic application on given "Click Days," starting at 8 AM, and such applications are processed on a first come–first served basis until available quotas of permits are exhausted. Exploiting discontinuities in "click time" to compare those just eligible for work permits to those not eligible, Pinotti (2017) finds that those eligible to work are significantly less likely to be arrested during the following year. The size of the effect is very large and remarkably similar, in relative terms, to that estimated by Mastrobuoni and Pinotti (2015)—a drop of more than 50 percent relative to the baseline crime rate—in spite of the fact that the two papers focus on very different populations; that is, former prison inmates and applicants for work permits. (Of course, the crime rate at the baseline differs widely between the two groups, being about ten times larger for former prison inmates than for applicants for a work permit.)

Other papers that rely on aggregate data confirm the conclusion that legal status to work has a strong effect on crime. In studying amnesties for undocumented immigrants across Italian regions, Fasani (2018) shows that they decrease crime rates, although the effect reported is much smaller than those estimated by Mastrobuoni and Pinotti (2015) and Pinotti (2017) using individual-level data. Outside Italy, Baker (2015) finds that the US Immigration Reform and Control Act (IRCA) of 1986—a generalized amnesty for irregular immigrants—caused a large decline in property crime in the counties relatively most affected by this policy. Relatedly, the end of the amnesty period after the passage of the 1986 legislation

coincided with a very large uptick in the number of arrests for economically motivated crimes (Freedman, Owens, and Bohn 2018).

Overall, this body of evidence confirms the prediction that acquiring legal status decreases immigrants' propensity to commit crime. Additional evidence suggests that the effect is driven by access to better economic opportunities in official labor markets. For instance, Mastrobuoni and Pinotti (2015) show that the decline in crime is driven by immigrants acquiring legal status in northern Italian regions, which experienced access to a dynamic labor market, while the effect is not significantly different from zero in southern Italian regions characterized by a much larger informal economy (and thus where legal status should not matter as much).

These findings also suggest some policy trade-offs. In the absence of perfect enforcement, low migration quotas and other types of legal movement restrictions may contribute to the formation of pools of irregular immigrants with limited access to legitimate earning opportunities, which in turn increases their propensity to engage in crime. In this static setting, generalized amnesties for irregular immigrants and other policies improving their labor market integration and earning opportunities may reduce their involvement into crime. In a dynamic setting, however, generalized amnesties of irregular immigrants may generate expectations of future legalizations, thus increasing migration pressures and arrival of undocumented migrants into the country.

## **Conclusion**

The 1931 Hoover Commission mentioned in the introduction, in its chapter on "Crime and the Foreign Born," states: "In proportion to their respective numbers the foreign-born commit considerably fewer crimes than the native-born" (National Commission on Law Observance and Enforcement 1931, p. 4). This conclusion was based on very simple statistical assessments. It has taken a surprisingly long time—especially when compared to the vast literature on the labor market effects of migration—for a systematic empirical analysis of the causal impact of immigration on crime to emerge in the economic literature.

The emerging evidence seems potentially inconsistent. On one side, in most countries—with the notable exception of the United States—immigrants exhibit a disproportionate involvement in criminal activity compared to natives, as measured by the relative incarceration rate of the two groups. In addition, certain kinds of immigrants, including young and less-educated men and those with irregular legal status, display a much higher propensity to commit crimes than those with documented status. These factors would seem to suggest a positive link between immigration and crime. On the other side, studies designed to measure the effect of immigration inflow effects on local crime rates do *not*, in general, find any detectable causal effect of immigration on local crime rates. For example, all previous studies relying on the shift-share instrumental variable approach estimate crime

elasticities close to zero in various countries, and we further confirm this result on new data across European countries and regions.

We can suggest several possible resolutions for this seemingly conflicting evidence. First, perhaps the observed high arrest and incarceration rates of immigrants stem from discrimination (preference-based or statistical) against immigrants on the part of police or judicial authorities, rather than actual higher crime rates, although we are not aware of any systematic review on this issue internationally. Second, perhaps immigrants have substituted for natives in some illegal criminal markets, which would be consistent with immigrants exhibiting higher offending rates than natives while, at the same time, immigration not leading to higher crime rates at the local level. Finally, the last (and perhaps simplest) explanation remains that the share of immigrants in the population—and among offenders—may still be in most cases too low to cause overall detectable changes in local crime rates. The share of immigrants in the global population is only 3.5 percent, and even in high-immigration countries in Europe and North America, it remains in most cases between 10 and 15 percent. These hypotheses are tentative, and further research is needed to reconcile the patterns emerging from individual-level and aggregate data on immigration and crime patterns.

Some recent work has sought to examine the persistence of the widespread perception that immigration worsens crime problems, in the face of the evidence of null or small effects of immigration on crime. A possible answer involves the role of (possibly biased) media in debates on immigration and crime. For example, Couttenier et al. (2023) study the lead-up to the passage of 2009 Swiss public referendum that banned construction of new minarets on mosques. The author used detailed data on actual crime along with information on newspaper coverage. They show a large upward distortion in media reporting of immigrant crime during the lead-up to the referendum. They estimate that the ban would have passed even in the absence of the heightened reporting, but the vote in favor would have been several percentage points lower. In Chile, where the stock of foreign-born residents with legally approved visas more than tripled from 2010 to 2017, Ajzenman, Dominguez, and Undurraga (2023) consider the effects on crime and on public perceptions of crime. As noticed earlier, using both a shift-share instrumental variable approach and an approach with two-way fixed effects at the municipality and year level, they find no causal effect of immigration on crime. However, they do find a large rise in public concerns about immigration and crime. In turn, these public concern within a given municipality seem linked to whether the inflow of immigrants in that area were less educated or with a non-European ethnicity. In addition, fears about immigrants and crime rose more sharply in municipalities with a larger number of local radio stations per capita, suggesting the possibility that competing media outlets were amplifying public fears.

■ We thank Eric Hurst, Enrico Moretti, Nina Pavcnik, Timothy Taylor, and Heidi Williams for useful comments when preparing this paper, and Pierre Robiglio and Andrea Giglio provided outstanding research assistance. Paolo Pinotti acknowledges financial support from ERC Consolidator Grant “CLEAN” (Contract #: 866181). Olivier Marie acknowledges financial support from the Netherlands Organization for Scientific Research (NWO 016.Vidi.185.049).

## References

- Abramitzky, Ran, Leah Platt Boustan, Elisa Jácome, Santiago Pérez, and Juan David Torres. 2023. “Law-Abiding Immigrants: The Incarceration Gap between Immigrants and the US-Born, 1850–2020.” NBER Working Paper 31440.
- Adão, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. “Shift-Share Designs: Theory and Inference.” *Quarterly Journal of Economics* 134 (4): 1949–2010.
- Ajzenman, Nicolás, Patricio Dominguez, and Raimundo Undurraga. 2023. “Immigration, Crime, and Crime (Mis)perceptions.” *American Economic Journal: Applied Economics* 15 (4): 142–76.
- Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. 2022. “Immigration and Redistribution.” *Review of Economic Studies* 90 (1): 1–39.
- Altonji, Joseph G., and David Card. 1991. “The Effects of Immigration on the Labor Market Outcomes of Less-Skilled Natives.” In *Immigration, Trade, and the Labor Market*, edited by John M. Abowd and Richard B. Freeman, 201–34. Chicago: University of Chicago Press.
- Andersen, Lars Højsgaard, Christian Dustmann, and Rasmus Landersø. 2019. “Lowering Welfare Benefits: Intended and Unintended Consequences for Migrants and Their Families.” Centre for Research and Analysis of Migration Discussion Paper 05/19.
- Baker, Scott R. 2015. “Effects of Immigrant Legalization on Crime.” *American Economic Review* 105 (5): 210–13.
- Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo: W. E. Upjohn Institute.
- Becker, Gary S. 1968. “Crime and Punishment: An Economic Approach.” *Journal of Political Economy* 76 (2): 169–217.
- Bell, Brian, Francesco Fasani, and Stephen Machin. 2013. “Crime and Immigration: Evidence from Large Immigrant Waves.” *Review of Economics and Statistics* 21 (3): 1278–90.
- Bianchi, Milo, Paolo Buonanno, and Paolo Pinotti. 2012. “Do Immigrants Cause Crime?” *Journal of the European Economic Association* 10 (6): 1318–47.
- Borjas, George J. 2003. “The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market.” *Quarterly Journal of Economics* 118 (4): 1335–74.
- Borusyak, Kirill, Peter Hull, and Xavier Jaravel. 2022. “Quasi-Experimental Shift-Share Research Designs.” *Review of Economic Studies* 89 (1): 181–213.
- Butcher, Kristin F., and Anne Morrison Piehl. 1998. “Cross-City Evidence on the Relationship between Immigration and Crime.” *Journal of Policy Analysis and Management* 17 (3): 457–93.
- Card, David. 2001. “Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration.” *Journal of Labor Economics* 19 (1): 22–64.
- Card, David. 2009. “Immigration and Inequality.” *American Economic Review* 99 (2): 1–21.
- Chalfin, Aaron. 2014. “What Is the Contribution of Mexican Immigration to U.S. Crime Rates? Evidence from Rainfall Shocks in Mexico.” *American Law and Economics Review* 16 (1): 220–68.
- Couttenier, Mathieu, Sophie Hatte, Mathias Thoenig, and Stephanos Vlachos. 2023. “Anti-Muslim Voting and Media Coverage of Immigrant Crimes.” *Review of Economics and Statistics*. [https://doi.org/10.1162/rest\\_a\\_01152](https://doi.org/10.1162/rest_a_01152).

- Dustmann, Christian, Uta Schönberg, and Jan Stuhler.** 2016. "The Impact of Immigration: Why Do Studies Reach Such Different Results?" *Journal of Economic Perspectives* 30 (4): 31–56.
- Ehrlich, Isaac.** 1996. "Crime, Punishment, and the Market for Offenses." *Journal of Economic Perspectives* 10 (1): 43–67.
- Eurostat.** 2023. "Crime and Criminal Justice Database." <https://ec.europa.eu/eurostat/web/crime/database> (accessed on December 11, 2023).
- EVS.** 2022. "European Values Study 2017: Integrated Dataset (EVS 2017)." GESIS Data Archive, Cologne. ZA7500 Data file Version 5.0.0, doi:10.4232/1.13897.
- Fasani, Francesco.** 2018. "Immigrant Crime and Legal Status: Evidence from Repeated Amnesty Programs." *Journal of Economic Geography* 18 (4): 887–914.
- Fasani, Francesco, Tommaso Frattini, and Luigi Minale.** 2021. "Lift the Ban? Initial Employment Restrictions and Refugee Labour Market Outcomes." *Journal of the European Economic Association* 19 (5): 2803–54.
- Fasani, Francesco, Giovanni Mastrobuoni, Emily G. Owens, and Paolo Pinotti.** 2019. *Does Immigration Increase Crime?* Cambridge, UK: Cambridge University Press.
- Freedman, Matthew, Emily Owens, and Sarah Bohn.** 2018. "Immigration, Employment Opportunities, and Criminal Behavior." *American Economic Journal: Economic Policy* 10 (2): 117–51.
- Friedberg, Rachel M.** 2001. "The Impact of Mass Migration on the Israeli Labor Market." *Quarterly Journal of Economics* 116 (4): 1373–1408.
- Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift.** 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review* 110 (8): 2586–2624.
- Grigorieff, Alexis, Christopher Roth, and Diego Ubfal.** 2020. "Does Information Change Attitudes Toward Immigrants?" *Demography* 57 (3): 1117–43.
- Haaland, Ingar, and Christopher Roth.** 2020. "Labor Market Concerns and Support for Immigration." *Journal of Public Economics* 191: 104256.
- Haerper, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., M. Lagos, P. Norris, E. Ponarin & B. Puranen et al., eds.** 2022. "World Values Survey: Round Seven—Country-Pooled Datafile." Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat. Version. 5.0.0, doi:10.14281/18241.20.
- Hirschi, Travis, and Michael Gottfredson.** 1983. "Age and the Explanation of Crime." *American Journal of Sociology* 89 (3): 552–84.
- Institute for Crime & Justice Policy Research (ICPR).** 2023. "World Prison Brief." Birkbeck, University of London. <https://www.prisonstudies.org/> (accessed on December 11, 2023).
- Italian Ministry of the Interior.** 2007. *Rapporto sulla criminalità in Italia: Analisi, Prevenzione, Contrasto*. Rome: Government of Italy.
- Jaeger, David, Joakim Ruist, and Jan Stuhler.** 2018. "Shift-Share Instruments and Dynamic Adjustments: The Case of Immigration." NBER Working Paper 24285.
- Knox, Dean, and Jonathan Mummolo.** 2020. "Making Inferences about Racial Disparities in Police Violence." *Proceedings of the National Academy of Sciences* 117 (3): 1261–62.
- Lochner, Lance, and Enrico Moretti.** 2004. "The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports." *American Economic Review* 94 (1): 155–89.
- Machin, Stephen, Olivier Marie, and Sunčica Vujčić.** 2011. "The Crime Reducing Effect of Education." *Economic Journal* 121 (552): 463–84.
- Marie, Olivier, and Paolo Pinotti.** 2024. "Replication data for: Title." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. <https://doi.org/10.3886/E143461V1>.
- Mastrobuoni, Giovanni, and Paolo Pinotti.** 2015. "Legal Status and the Criminal Activity of Immigrants." *American Economic Journal: Applied Economics* 7 (2): 175–206.
- Mayda, Anna Maria.** 2006. "Who Is against Immigration? A Cross-Country Investigation of Individual Attitudes toward Immigrants." *Review of Economics and Statistics* 88 (3): 510–30.
- National Commission on Law Observance and Enforcement.** 1931. *Report on Crime and the Foreign Born*. Washington, DC: US Government Printing Office.
- OECD.** 2018. "Settling in 2018 : indicators of immigrant integration." OECD Publishing. <https://data.europa.eu/doi/10.1787/9789264307216-en>.
- OECD.** 2023. "Stocks of foreign-born population in OECD countries (indicator)." doi: 10.1787/b1fc67fa-en (accessed on December 11, 2023).
- Ousey, Graham C., and Charis E. Kubrin.** 2009. "Exploring the Connection between Immigration and Violent Crime Rates in U.S. Cities, 1980–2000." *Social Problems* 56 (3): 447–73.



- Peri, Giovanni.** 2016. "Immigrants, Productivity, and Labor Markets." *Journal of Economic Perspectives* 30 (4): 3–30.
- Pinotti, Paolo.** 2017. "Clicking on Heaven's Door: The Effect of Immigrant Legalization on Crime." *American Economic Review* 107 (1): 138–68.
- Piopiunik, Marc, and Jens Ruhose.** 2017. "Immigration, Regional Conditions, and Crime: Evidence from an Allocation Policy in Germany." *European Economic Review* 92: 258–82.
- Quetelet, Adolphe.** 2003. "Research on the Propensity for Crime at Different Ages." Translated by S. F. Sylveski. In *Crime: Critical Concepts in Sociology*, edited by Philip Bean, 119–35. London: Routledge, (Orig. pub. 1831).
- Reid, Lesley Williams, Harald E. Weiss, and Robert M. Adelman, and Charles Jaret.** 2005. "The Immigration–Crime Relationship: Evidence across US Metropolitan Areas." *Social Science Research* 34 (4): 757–80.
- Spenkuch, Jörg L.** 2014. "Understanding the Impact of Immigration on Crime." *American Law and Economics Review* 16 (1): 177–219.
- Tuttle, Cody.** 2019. "Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums." Unpublished.
- United Nations.** 2023a. "Global Migration Database." <https://www.un.org/development/desa/pd/data/global-migration-database> (accessed on December 11, 2023).
- United Nations.** 2023b. "Global Study on Homicide." <https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide.html> (accessed on December 11, 2023).
- Wadsworth, Tim. 2010. "Is Immigration Responsible for the Crime Drop? An Assessment of the Influence of Immigration on Changes in Violent Crime between 1990 and 2000." *Social Science Quarterly* 91 (2): 531–53.



# Care Provision and the Boundaries of Production

Nancy Folbre

**E**conomists generally treat care for others—including expenditures of time and money by individuals, families, communities, states, and businesses—as a form of redistribution. Yet care provision can be conceptualized more broadly as the production, development, and maintenance of human capabilities whose value extends far beyond individual rates of return in the labor market. Capabilities cannot be reduced to what economists typically refer to as skill; they include physical and mental health and the ability to collaborate with others to solve coordination problems that lie beyond the scope of market exchange. Collective commitments to the intrinsic value of human capabilities are crucial to any vision of sustainable economic development.

Most economic textbooks picture households as units of exchange within the circular flow of market income—supplying labor to the market in return for income to purchase goods and services rather than contributing to “output.” While the conventional household economics developed by Gary Becker and Jacob Mincer partially challenges this assumption, it focuses primarily on the production of utility or subjective satisfaction and increases in future market earnings, rather than overall contributions to the economy. While recent research on household bargaining calls attention to factors influencing distribution within households, it says little regarding the social contributions of unpaid care.

A tendency to consider productive activities *internal* to households has been accompanied by a tendency to overlook productive contributions *external* to markets.

■ Nancy Folbre is Professor Emerita of Economics, University of Massachusetts Amherst, Amherst, Massachusetts. Her email address is [folbre@econ.umass.edu](mailto:folbre@econ.umass.edu).

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.1.201>.

Most discussion of externalities focuses on the natural environment, ranging from the pollination produced by a beekeeper's honey business (a positive externality) to the effects of greenhouse gas emissions on the global climate (a negative externality). Yet successful care provision can improve the social climate and contribute to what is sometimes termed social capital. Conversely, shortfalls in care provision can increase social costs such as crime, drug addiction, suicide, and more generally, poor mental and physical health.

Attention to care provision, as conceptualized above, extends the boundaries of economic thinking in three important ways. This essay (1) highlights the need to consider the productive contributions—including fiscal and social externalities—of unpaid work and other nonmarket transfers; (2) challenges the assumption that economic rewards for unpaid and underpaid care work can be fully explained either by the subjective satisfaction they provide or their market value-added; and (3) underscores the impact of social institutions on the economic bargaining power of care providers.

All three points have precedents in received economic theory. Applied in concert with recent research quantifying inputs into care provision, they challenge the assumptions of traditional macroeconomics and neoclassical theories of wage determination. They also reveal the impact of distributional conflict and collective power on social institutions, with important implications for public policy.

## **Measuring Care Provision**

Empirical exploration requires clear and consistent definitions, which may, in turn, be altered and improved by further research. Amartya Sen (1999) provides a philosophical defense of the intrinsic value of capabilities, consistent with a broad definition of human capital that encompasses social as well as individual rates of return. Empirical researchers have reached for ways to operationalize these concepts. The Human Development Index (HDI) calculated by the United Nations provides an early example of an effort to look beyond a market metric, including averages of life expectancy and years of schooling alongside per capita income as an indicator of successful development. Such examples have encouraged efforts to categorize human capabilities that go beyond measures of health and education (Nussbaum 2011). They have also increased interest in measuring inputs into human capabilities that reach beyond traditional concepts of investment in education and job market experience.

### **Definitions**

With the advent of nationally representative time use surveys, attention has turned to measures of time devoted to unpaid work, defined as activities that someone else could, in principle, be paid to perform. Women devote significantly more time to unpaid work than men, even when employed full-time. Researchers typically distinguish between “direct care” that involves face-to-face interaction (such as feeding or bathing a child) and “indirect care” such as housework that maintains the care

environment with attention to person-specific needs (Folbre 2012). Some categories of time-use fall between these two categories, such as time devoted to the necessary supervision of young children, which may be combined with housework or leisure, but nonetheless constrains opportunities for employment outside the home.

Unpaid care work and private expenditures on care provision are obviously related and at least partially substitutable. Care provision, however, is not limited to labor services. People cannot acquire education and skills unless they are fed, housed, and literally provided for. Parental expenditures on meeting children's basic needs are just as necessary as direct inputs of parental time. Researchers sometimes treat the official poverty line as a useful threshold, but expenditures on children vary dramatically by family income (National Academies of Sciences, Engineering, and Medicine 2019). The term "investments in children" is often used loosely because it is difficult to draw a clear line between investment and consumption. Indeed, attention to care provision suggests that these categories can overlap, an issue that invites more conceptual and empirical attention.

Efforts to categorize paid care inputs parallel the distinctions applied to unpaid care work. Care occupations such as childcare, eldercare, nursing, and teaching typically involve direct care: face-to-face interactions that typically include concern for the wellbeing of care recipients. Employment in care service industries such as health, education, and social services involves indirect care, distinguished less by the labor process than by the type of output—improvements in capabilities that create significant positive social externalities. Not surprisingly, most care occupations are situated within specific care service industries. The wages that employees earn are often interpreted as a measure of their value-added. Where both intrinsic motivation and externalities come into play, this interpretation falls short.

The simplest approach to measuring public contributions to care provision is to focus on expenditures on health and education, which, when made by individuals, are often described as investments in human capital. In principle, forms of social assistance aimed at reducing poverty and privation could also be interpreted as contributions to care provision, and their effects on human capabilities and the social climate deserve serious attention.

### **Dimensions**

Economic analysis does have boundaries; time and money are obviously not the only inputs into human capabilities. Many other factors, including commitment, emotional intelligence, and continuity of care are also vital. Still, the quantitative dimensions of measurable resources devoted to care provision command attention. The American Time Use Survey (<https://www.bls.gov/tus/>) shows that average amount of time per day that adults devoted to housework activities, shopping, and direct care of household and nonhousehold members is about 3.3 hours, not much lower than the 3.5 hours devoted to paid employment. Add in the 0.1 average hours spent in volunteering (a kind of community care) and 0.4 hours spent in education (self-investment), and average care time amounts to about 3.7 hours per day for all adults and about 4.5 hours per day for women.

Child care is a particularly demanding task that can seldom be postponed. Women living in a household with a child under six spent on average of about 2.8 hours a day in primary activities such as feeding, bathing, and changing diapers, and about 7 hours a day in supervisory or “secondary” time with children “in their care.” Men are less likely than women to live in households with children under six but those who do average about 1.6 hours a day in primary childcare and about 4.8 in supervisory or “secondary” time.

The US Department of Agriculture occasionally provides estimates of annual family expenditures on children up to age 18; in 2015 these amounted to a range of \$16,000 to \$18,000 per year (in 2023 dollars) per child in married couple, two-child, middle-income families, not including any estimate of the value of parental time. If parental time is assigned a market value equivalent to state-level minimum wages, the average value of parental resources devoted to children more than doubles for all children under age 15 in such families (Folbre and Gautham forthcoming).

The care needs of adults with disabilities (including the frailties of old age) is more intermittent and difficult to predict, often varying considerably over the course of a week or a year. Family members, particularly women, often step up. An estimated 38 million people provided direct care to adults with chronic, disabling, or serious health conditions in 2021. The cost of purchasing market substitutes for this time has been estimated at about \$600 billion, or about 14 percent of total health care spending in that year (Reinhard et al. 2019; American Medical Association 2023). Family, friend, and neighbor care also represents an indispensable form of informal insurance against the risks of temporary illnesses and accidents.

The quantitative dimensions of paid care services are also large, and reveal significant segregation by gender: In 2022, about 22 percent of all employees, and 36 percent of all women employees, were employed in education and health services (including social assistance) (see the Bureau of Labor Statistics website at <http://bls.gov/cps/cpsaat18.htm>). Broken out by occupation in the same year, about 18 percent of all employees and 29 percent of all women employees worked in the major occupational categories of community and social services, education, training, library occupations, health care practitioners, health care support, and childcare (<http://bls.gov/cps/cpsaat11.htm>).

Public sector contributions to care provision are huge: mandated expenditures on Social Security and major health care programs alone were projected to account for 46 percent of all federal outlays in 2023 (Congressional Budget Office 2023). In 2020, state and local governments spent about 39 percent of their budgets on health care and elementary, secondary, and higher education (Urban Institute 2023).

Public spending has a significant intergenerational component, as occurs when the working age population is taxed to help finance transfers to both the young and the old. The National Transfer Accounts project (<https://www.ntaccounts.org/web/nta/show/>) examines the size and direction of both private and public intergenerational transfers in the United States as well as many other countries. At least one study estimates the overall size of the “care sector”—including unpaid care, paid care, and public spending—on the state level (Duffy, Albelda, and Hammonds 2013).

## **Implications for Macroeconomics**

Economists generally measure value by multiplying prices by quantities, but this approach falls short in a sector where many inputs are either unpaid, unpriced, or priced simply by input costs, as in most government provision. For many years, the US National Income and Product Accounts (NIPA), like the International System of National Accounts (SNA), has placed unpaid services provided in the home “outside the production boundary” (BEA 2019). This boundary has increasingly come into question, undermined by growing criticism backed by the evidence of time-use surveys. The United States now provides a “satellite” national income account that includes imputations of the market value of nonmarket work. These imputations represent a step in the right direction, but are based entirely on valuation of inputs, with no consideration of the actual value-added of person-specific services. Consideration of the fiscal and social externalities of care provision undermines basic assumptions of macroeconomics and public finance, even as it generates new empirical challenges.

### **Imputing the Value of Unpaid Work**

Valuation of nonmarket services raises thorny issues. So-called “replacement cost” estimates are based on a hypothetical question: “What would it cost to hire someone to provide service of the same quality?” Unfortunately, some circularity comes into play. The extensive supply of unpriced labor lowers the market price of many of the paid services available as substitutes for it. Also, quality is difficult to assess: Skilled and specialized tradespeople are probably more productive than the average householder at some tasks, but person-specific skills make unpaid care of family members particularly valuable. Paid and unpaid services are not perfectly substitutable, and households may have basic thresholds of need for both.

By comparison, opportunity cost calculations ask: “What am I giving up by engaging in this activity?” The answer is often based on forgone earnings. However, additional hours of employment are not always available. Moreover, personal choices are affected by unobservable preferences as well as relative prices, and there is no reason to believe that a college professor is any better at changing diapers than a high school graduate, even if her time is more costly. While opportunity cost valuation is relevant to individual decisions, it says little about the level of benefits provided to others.

The replacement cost method is most widely applied in satellite national income accounts, which insert imputations of the value of unpaid household services into measures of extended gross domestic product. The US Bureau of Economic Analysis began publishing such an account in 2000 (Landefeld and McCulla 2000). The latest estimate applies an average household workers’ wage (\$12.71 per hour) to hours as measured in the American Time Use Survey and assigns a value to the services of consumer durables based on returns to household financial assets. By this account, household production accounted for about 25 percent of total US output in 2020 (Bridgman, Craig, and Kanal 2022).

Whatever its limitations, replacement cost accounting dramatizes macroeconomic dynamics that had previously received only passing attention. For instance, conventional estimates of economic growth are overstated when they do not account for declines in hours of unpaid work associated with increases in women's paid employment. Recalculated growth rates of output for the late nineteenth- and early twentieth-century US economy based on imputations from Census labor force data illustrate this effect (Folbre and Wagman 1993; Wagman and Folbre 1996). According to the most recent Bureau of Economic Analysis estimate, the average rate of growth of nominal GDP from 1965 to 2020 drops from 6.3 to 6.1 percent even when the value of nonmarket production is imputed using conservative assumptions regarding replacement cost (Bridgman, Craig, and Kanal 2022). Time-use data also show that increases in unpaid household services can slightly buffer the effects of unemployment on household living standards, a finding consistent with more detailed analysis of time use during the Great Recession (Aguiar, Hurst, and Karabarbounis 2013).

Valuation of unpaid work has implications for measurement of income and consumption on the household as well as the national level. Households with a full-time homemaker enjoy more "extended income" (market income plus the imputed value of unpaid household services) than two-earner households that are similar in other respects (Folbre, Murray-Close, and Suh 2018; Folbre et al. 2013). Retirees compensate for lower expenditures by engaging in more unpaid household work (Aguiar and Hurst 2005; Folbre, Reimers, and Yoon 2009). Because households vary far less in time devoted to unpaid work than in market earnings, extended income is generally more equally distributed than market income (Frazis and Stewart 2011). Because the relative amount of time devoted to unpaid work has decreased since the 1960s, its equalizing effects have probably also diminished. Empirical research on tradeoffs between money expenditures and unpaid work time has long been hampered by the lack of household surveys covering both.

Input-based valuation is also applied to the output of workers in publicly financed care services of health and education, because the value of the human capabilities they develop or maintain cannot be easily assigned a market value. As elsewhere in the public sector, the value of output is simply set equal to outlays. Because the actual product is not measured, productivity cannot be measured either. Growth in the relative importance of care services reduces the measured rate of productivity growth in the economy as a whole, even when these services are generating important nonmonetary benefits such as increased life expectancy or higher rates of technical innovation. Even the output of private firms selling care services is underestimated, because their revenue is based on what customers are willing and able to pay, without regard for positive externalities.

### **Intergenerational Transfers**

Attention to nonmarket transfers expands the scope of intergenerational transfers far beyond bequests of wealth. The relatively new National Transfer Accounts project collects data on intrafamily transfers of income and wealth and examines



differences between taxes paid and public benefits received. Imputations of the value of unpaid household services have been added to these accounts for a number of countries (Gál, Vanhuyse, and Vargha 2018). Extensive calculations for the United States show that public transfers to the elderly through Social Security and Medicare are counterbalanced by extensive family expenditures on the younger generation (Lee et al. 2017).

The contrasting directions of these transfers raise provocative questions concerning private contributions to public fiscal benefits. The US Social Security system taxes employees to finance health and retirement benefits for retirees. These benefits, linked to the earnings history of recipients and their spouses, offer no credit for the time and money put into raising future taxpayers. As a result, the net benefits of Social Security are significantly higher for individuals who devote relatively little time or money to children, even though single-earner married couples reap higher net benefits than dual-earner or single-parent households. US public policies socialize the economic benefits of children far more extensively than the costs (Folbre and Wolf 2013).

Parents—defined in terms of contribution to childrearing rather than biology—create a significant fiscal externality by raising the next generation of taxpayers (Wolf et al. 2011). While empirical research has not yet distinguished between the total contributions of mothers and fathers or impact of greater life expectancy (and thus greater take-up of Social Security and Medicare) among women, mothers clearly make larger fiscal contributions through this channel than either childless individuals or noncustodial parents who fail to provide much support for their biological children.

The need to provide for an elderly population that is growing proportionately larger as a result of declining fertility and rising life expectancy accentuates the importance of increasing the productivity of the younger generation. A growing research literature documents long-run payoffs to investments in children's health and education, both for children themselves and for society as a whole (Currie 2020). Spending on children, whether public or private, can be considered a special category of investment that is key to integrating demography into macroeconomic theory. Likewise, improvements in life expectancy and reductions in chronic illness and disability represent valuable contributions to the capabilities of adults.

New accounting frameworks do not and probably cannot provide precise estimates of the market value of nonmarket transfers of time and money. However, they reveal the need for new macroeconomic models that consider the impact of changes in the relative size of these transfers and the multiplier effects of high quality care provision. They also raise provocative questions regarding the distribution of social costs and benefits.

## **Implications for Microeconomics**

Economists typically interpret individual decisions as interactions between fixed preferences, budget constraints, and market prices. Care provision demands

a more complicated story attentive to the impact of changing cultural norms and social contexts on individual preferences. While care provision is sometimes equated with caregiving, the intrinsic motivation implied by “caregiving” cannot be taken for granted. Altruistic preferences are not God-given or programmed in human DNA. They grow out of our history as a social species that encourages responsibilities for the care of others, especially dependents.

Women have historically devoted more time and energy than men to the direct care of others and have generally been held to higher standards of moral obligation to do so (Folbre 2009). However, the insight that caring preferences partly reflect moral values does not imply that virtue is its own reward. Nor does it imply that moral values are impervious to pressure. Indeed, economic reasoning itself suggests that people respond to increases in relative costs. Economists faithful to the principle of revealed preference assume that greater “psychic income” compensates for lower money income, but there is no direct evidence that it does. Women may well enjoy family commitments and flexibility in employment more than men do (Fuchs 1988; Goldin 2021), but this does not nullify the economic risks of care provision. Some people may make choices that maximize their lifetime utility; others may make incorrect assumptions or learn that they are locked into costly choices. In the real world where people often experience regret, women who make costly choices are not necessarily happier as a result.

Care provision is costly for individuals for several reasons. It often entails commitments to dependents—or to future generations—with no clear guarantees of reciprocity or payback. If concern for the well-being of others is distributed unequally, it reduces the bargaining power of care providers, who are reluctant to threaten withdrawal of their services. Individual value-added is difficult to measure, much less to capture or reward. Human capabilities have public good dimensions, making it easy to free ride on the contributions of others. Whether care is provided inside or outside the market, its pecuniary rewards are seldom aligned with its social benefits.

### **Cultural Norms and Endogeneity**

Cultural norms influence personal preferences, in part because conformity to them is likely to yield subjective satisfaction. Yet gender norms are vulnerable to collective renegotiation and change over time. In 1977, 65 percent of US respondents to the General Social Survey (<https://gss.norc.org/>) agreed with the statement: “It is better for everyone if the man is the achiever outside the home and the woman takes care of home and family.” By 2018, only 25 percent agreed. It is worth noting that even in the later year, survey administrators felt comfortable with wording implying that taking care of home and family is not an achievement.

In more subtle forms, traditional gender norms remain influential. Evidence suggests that they limit women’s bargaining power over leisure time in the household (Bittman et al. 2003; Bertrand, Kamenica, and Pan 2015) and also channel women into less well-paying jobs (Fortin 2008). Some preferences consonant with

personality traits have clear economic consequences: “Machiavellianism” is associated with higher earnings, empathy with lower earnings (Bowles, Gintis, and Osborne 2001; Kamas and Preston 2020). The heterosexual dating and marriage market tends to reward men—but not women—who signal a strong desire for high earnings (Badgett and Folbre 2003; Burstyn, Fujiwara, and Pallais 2017). The same investments that enhance women’s future earnings may reduce their options for family formation.

Care provision itself may modify preferences, leading to emotional attachments that reinforce initial patterns of gender specialization. Much of the economic research on loss aversion focuses on attitudes toward material possessions such as coffee cups or money (Kahneman, Knetsch, and Thaler 1990). Emotional attachments are even more powerful and consequential. As one grandmother who became involved in caring for her grandson put it, “I didn’t expect this and I didn’t want it, but my heart’s involved now” (as quoted in Associated Press 2002). Paid care providers often describe a similar endogeneity: “I love them. That’s all, you can’t help it” (Stone 2000, p. 99).

Patriarchal institutions that once strictly limited women’s choices channeled them into roles that virtually guaranteed an ample supply of care. Emerging cultural norms that elevate the personal pursuit of economic self-interest for women as well as for men veer to the opposite extreme. The cultural dissonance festering around the gender binary betrays a certain anxiety: If women are more likely to adopt traditionally masculine priorities than the other way around, we might expect the supply of care to decline. Indeed, men may resist embracing traditionally feminine priorities partly because they correctly perceive them to be costly.

### **Care Constraints and Penalties**

Constraints as well as motivations shape care provision. While it is important to ask why people make commitments to provide care, we should also ask how the costs of these commitments are distributed. A labor market based on exchanges between employers and employees offers wages and salaries based on perceptions of individual value-added and relative bargaining power, not larger contributions to care provision. Employers have little stake in rewarding efforts that do not contribute to measurable institutional goals, unless these efforts improve their ability to retain their most valuable employees.

Mothers pay a hefty penalty in lower lifetime earnings for each child, often resulting in greater vulnerability to poverty for themselves and their children (Kleven et al. 2019). A recent report published by the Urban Institute in conjunction with the Women’s Bureau of the US Department of Labor estimates women’s lost lifetime earnings resulting from the care of family members (predominantly children) at \$237,000 (in 2021 dollars), or 15 percent of women’s projected lifetime earnings. These costs vary significantly by race, level of education, and family structure (Johnson, Smith, and Butrica 2023). Poorly educated, low-earning women, disproportionately Black and Hispanic, are particularly likely to become a sole source of support for dependent family members.

Costs are not limited to unpaid care provision. Workers in care service industries and occupations suffer a wage penalty, after controlling for education and many other factors (including individual fixed effects) (Folbre, Gautham, and Smith 2022; Budig, Hodges, and England 2019; England, Budig, and Folbre 2002). Such penalties affect men as well as women, but women are more heavily affected, due to their concentration in these jobs. Women of color and immigrants are over-represented in the lowest paid care occupations—childcare and long-term care for the elderly and disabled.

Some research parses the effect of specific inputs, such as more highly rated teachers, on specific types of outputs, such as higher earnings (Chetty, Friedman, and Rockoff 2014). However, teachers are obviously not rewarded for their individual contributions to their students' lifetime skills, just as parents are not rewarded for their value-added to their children. Likewise, nurses and doctors are not rewarded for their individual contributions to patients' lifetime health, and social workers are not rewarded for their individual contributions to social welfare. Industry wage differentials' net of personal characteristics partly reflect cross-industry differences in ability to capture value-added: Unlike employees in care services, employees in business services generate measurable and significant revenues for their employers, are more easily paid for performance, and earn significantly more (Folbre, Gautham, and Smith 2022).

## **Social Institutions and Collective Power**

Partly as a result of the limits to market exchange, societies rely on implicit and explicit institutional arrangements to govern care provision. Implicit and explicit social contracts do not evolve simply because they solve coordination problems or enhance efficiency. They typically reflect the disparate power of groups defined by dimensions of group identity such as gender, age, sexuality, race, class, and citizenship. Feminist research traces the emergence and evolution of patriarchal contracts that have long shifted a disproportionate share of the costs of producing, developing, and maintaining human capabilities onto women and other disempowered groups (Folbre 2021).

For much of US history, women were denied basic property rights and political representation. The impact of state and federal laws passed between 1869 and 1920 giving women the right to vote provides a powerful example of the ways in which shifts in political power can alter investments in human capabilities. Not surprisingly, those who specialized in care provision tended to advocate more public support for it. Women's electoral mobilization led to significant increases in local spending on public health, including hygiene campaigns that tackled infectious disease and reduced child mortality by 8–15 percent (Miller 2008).

Still, public policies often enforced disparate responsibilities based on gender. In the first half of the twentieth century, many school districts required women teachers to resign upon marriage; during the Great Depression of the 1930s, federal

hiring practices forced the resignation of employees married to other federal employees, primarily affecting wives (Goldin 1990). Until the 1960s, patterns of discrimination and segregation from Jim Crow legislation restricted many Black and Hispanic women to jobs in domestic service. Both public and private employers openly discriminated against women before the 1963 Equal Pay Act and the 1964 Civil Rights Act, which, both gradually reduced gender inequality in pay (Bailey, Helgerman, and Stuart 2023).

Forced specialization also took the form of restrictions on women's reproductive rights. Forms of contraception easily deployed by women (such as diaphragms) faced far greater social disapproval and legal sanction in the early twentieth century than those traditionally used by men (such as condoms). Differences in the timing of state legislation legalizing birth control pills reveal highly significant effects on women's earnings (Bailey, Hershbein, and Miller 2012). Research on the timing of abortion legalization across states dating from the US Supreme Court's initial *Roe v. Wade* (410 US 113 [1973]) decision shows that it reduced maternal mortality and increased the education, employment, and earnings of Black women in particular (Myers and Welch 2021). The effects of the Supreme Court reversal in *Dobbs v. Jackson Women's Health Organization* (597 US \_\_\_ [2022]) remain to be seen.

Despite an overall pattern of advances in both civil and reproductive rights, steady gains in the employment rates of women starting in the 1970s began to plateau in the 1990s, prompting attention to the competing demands of paid work and family care. At the top of labor market, many employers prefer professionals and managers who are unconstrained by family responsibilities and able to put in exceptionally long hours on the job (Goldin 2021). Women at the bottom end in low-wage service jobs often face the opposite problem: just-in-time staffing with unpredictable hours making assistance with childcare difficult to find (Lambert, Henly, and Kim 2019). Black and Hispanic women are overrepresented among single mothers living in poverty, who face particularly acute time binds (Albelda 2011). Covert employer discrimination against mothers of young children is still evident (Ishizuka 2021).

Women may choose their own commitments to provide unpaid care, but they have little collective voice in determining the costs and risks incurred. Many biological fathers default on economic obligations to their offspring: In 2017, less than one-half of single custodial parents had a legal or informal child support agreement, and less than one-half of those received the full amount they were owed (Grall 2020). Public policies such as subsidized care provision for children and adults experiencing illness or disability, paid family leaves from employment, paid sick leaves, scheduling flexibility, and public support in the form of family allowances or tax credits generally reduce the costs of private care provision. Such policies, adopted far more extensively in northwestern Europe than in the United States, help explain why women's employment rates have grown less rapidly here in recent decades (Blau and Kahn 2017). Pro-family policies can boost GDP by increasing maternal labor supply to the market; implications for the quality of unpaid care provision reach even further.

Resistance to policies that offer important public benefits for caregiving derives in part from multidimensional inequalities. Many highly paid women at the top of

the corporate ladder have sufficient bargaining power to command fairly generous paid family and medical leave from their employers. In 2017–2018, women in the top quartile of earnings were 3.5 times more likely to have rights to paid leave than those in the lowest quartile; Asian, Black, and Hispanic women workers were all significantly less likely than White non-Hispanic women to have access (Goodman et al. 2022). While paid family leave is often described simply as a convenience, it is associated with improved health for both mothers and children (Bullinger 2019). Men also benefit from improved opportunities for direct care, though they are less likely than women to take advantage of paid family leave (Petts, Knoester, and Li 2020).

In recent years, the flow of low-wage migrants to the United States has helped restrain increases in the costs of outsourced services such as housecleaning, child-care, elder care, prepared foods, and restaurant meals. Such outsourcing makes it easier for high-earner households to accommodate their own care needs (Romero 2018). In metropolitan areas with large numbers of low-wage migrants, women in the top quartile of earnings have enjoyed a small but significant boost in their average hours of employment and their earnings (Cortés and Tessada 2011). Unfortunately, market solutions at hand for the affluent also weaken political support for public care provision.

Some economists argue that state interventions “mimic the agreements that would occur if children were capable of arranging for their care” (Becker and Murphy 1988). Little evidence supports this claim. As findings from the National Transfer Accounts (discussed earlier) show, the older US population commands more generous public transfers than the young and is far less susceptible to poverty. Individuals under 18 cannot vote. Racial/ethnic differences are salient. Less than 50 percent of all those under 18 are non-Hispanic Whites, compared to 77 percent of those over 65 (Vespa, Medina, and Armstrong 2018). In 2019, over 70 percent of children living in poverty were children of color, defined as members of any racial category other than White (Children’s Defense Fund 2021). A political coalition that crosses generational, gender, and racial/ethnic lines to successfully support broad-based care policies has yet to emerge.

## **Reframing Care Policy**

While many vectors of distributional conflict create obstacles to the reform of care provision in the United States, conventional economic thinking that disregards its productive contributions must take some of the blame.

### **Measures of Success**

Market income is a misleading indicator of economic success. Imputing the replacement cost value of labor inputs into unpaid services improves the big picture, but nonetheless understates the true value of productive activities. Indicators of health, education, subjective well-being, and other capabilities could be included

on a dashboard of multiple social indicators alongside measures of the depreciation of natural assets and threats to ecological sustainability (Stiglitz, Fitoussi, and Durand 2019). Our national data infrastructure requires expansion. Existing surveys fail to provide adequate measures of inputs into care provision, much less their impact on social and economic outcomes (Folbre, Fremstad, and Gonalons-Pons 2023).

Considerable research documents significant economic payoffs to reductions in child poverty and improvements in early childhood education and health (National Academies of Science, Engineering, and Medicine 2019; Aizer, Hoynes, and Lleras-Muney 2022). These payoffs extend to improvements in adult health and wellbeing. Poverty, inequality, and economic stress reduce people's ability to care for themselves and for others, often with devastating consequences (Wilkinson and Pickett 2011; Case and Deaton 2020). The benefits of these programs accumulate over decades, but economists often focus on the short- and medium-run. For instance, the US Congressional Budget Office, responsible for cost-benefit analysis of the effects of proposed legislation, often restricts itself to a ten-year window for calculation (Aizer, Hoynes, and Lleras-Muney 2022). A longer time frame is needed. Discount rates, typically explained simply as time preference, can also be interpreted as devaluation of the welfare of future generations.

State care policies, including abortion rights, access to paid family leave, child support enforcement, and Medicaid support for home and community-based care, now vary widely across the country. The idiosyncratic terminology and budget structure of these programs makes it difficult to provide a comprehensive picture of their results. Greater efforts to monitor and compare care policies on the state level could both aid empirical research and help motivate changes in federal programs.

### **The Work Word**

Seemingly academic debates over the definition of "work" can have momentous consequences. In 2022, the proposed extension of an expanded Child Tax Credit that had dramatically reduced child poverty met vehement opposition from critics who argued that it should include a (paid) work requirement (as reported by Konish 2022). Failure to acknowledge the economic value of unpaid work has long distorted policy assessment. For instance, the Personal Responsibility and Work Opportunity Act of 1994, commonly known as the "welfare reform act," successfully increased paid employment but also reduced the time available for unpaid childcare, often increasing out-of-pocket expenses for welfare recipients. By one calculation, single mothers in the bottom half of the consumption distribution who valued their unpaid work time at more than \$3 per hour were worse off after the reforms than before (Meyer and Sullivan 2008, p. 2237).

The Earned Income Tax Credit (EITC), currently the largest US cash transfer program, requires recipients to earn significant income to gain eligibility. Yet many engage in unpaid work that is essentially the same as what they could be paid to do. For example, two mothers who care for their own infants full-time without pay do not qualify for the tax credit, but if they traded infants for eight

hours a day, charged one another the same for the service, and met administrative requirements such as paying Social Security taxes, they could both qualify for the maximum credit.

Even when the work of caring for kin is offered financial support, such work is often discounted. The level of assistance offered by current welfare programs under Temporary Assistance to Needy Families (TANF) is lower than the stipends offered to foster families, and some states offer lower stipends for kin than non-kin foster parents (Doyle and Peters 2007). In Massachusetts, as in several other states, adult sons and daughters, but not spouses, qualify to be paid for the care they provide through Medicaid-financed home and community-based assistance.

Public payment for care can crowd out family provision, but it can also crowd it in by reducing constraints that make it difficult for many people to provide the level of family care that they prefer. Both publicly provided and purchased care services can change the composition rather than the level of unpaid care. For instance, utilization of paid child care services reduces the time parents spend supervising children, but has little effect on time parents devote to interactive care (Suh and Folbre 2023). Families that make use of paid care services for help with routine physical care of a person experiencing disability have more time for care management and emotional support. Public support for care provision honors its contribution. By contrast, treating it as the result of individual preferences implies that people who do not enjoy providing family care should feel free to opt out.

### **Synergies**

Many proposed public policies, including publicly funded childcare and requirements for paid family leave and sick leave, are treated as mere amenities. They should be interpreted instead as support for economically valuable care provision. The availability and price of paid childcare is a key determinant of women's contributions to family income and also helps parents juggle demanding schedules (Ruppaner et al. 2021). High quality early childhood education makes tangible contributions to the health, educational attainment, good behavior, and earnings of those who receive it (Bartik 2014). Access to subsidized kindergarten can increase the time that mothers supply to the informal care of other family members needing assistance (Chari and Valli 2021). California's paid family leave policy—often simply interpreted as a benefit for parents—has been linked to an 11 percent decline in elderly nursing home utilization (Arora and Wolf 2018).

Hospital discharge procedures that explicitly integrate unpaid care providers reduce costs by shortening hospitalization, reducing readmission, and lowering expenditures on post-discharge care (Rodakowski et al. 2017). Because many health care institutions take this unpaid labor supply for granted, many states have recently passed the Caregiver Advise, Record, Enable (CARE) Act, which nudges hospitals to increase coordination and support for needed knowledge and skills (Hudson et al. 2020). In recent years, extensive redirection of Medicaid funds away from nursing homes toward home and community-based care has increased demands on family members to coordinate and bridge gaps (Basu et al. 2022). Proposed changes



to Social Security could give unpaid care providers credits toward retirement benefits of their own (Munnell and Eschtruth 2018).

### **Fragmentation**

The provision of care is highly fragmented and inefficient. Despite evidence of high administrative costs and poor coverage, efforts for additional reform of the health care system have gained little traction since the passage of the Patient Protection and Affordable Care Act of 2010. Neither public nor private health insurance adequately covers long-term care for chronic conditions such as dementia. As the size of the frail elderly population increases, the need for public long-term care insurance will intensify. In recent years the demand for childcare services has far exceeded the available supply (US Treasury 2021). Staff shortages and high turnover rates threaten the quality of care in both nursing homes and home and community-based care.

Incentives for improvement of care services are poorly aligned. Employer-based family and sick leave policies place much of the cost of private health care on the most generous employers. The quality of nursing home care declines as unemployment declines, partly because institutions relying on Medicaid reimbursements cannot afford to pay the higher wages necessary to retain workers (Huang and Bowlblis 2019). Efforts to develop a rating system to improve consumer choice of nursing homes have overlooked important concerns, such as needs for close physical proximity to family members. Expansion of home and community-based services could improve health outcomes, but third-party payers of hospital and physician services have little incentive to divert medical savings to reward services currently funded by private payers or Medicaid (Doty 2017, p. 117).

A more unified and universal system of care provision, including more support for unpaid care through a basic income payment, family allowance, or refundable tax credit, could help bridge political differences. In the United States, the activist group Caring Across Generations (<https://caringacross.org/>) advocates for broader and more generous care policies. Nordic countries, as well as Canada, provide a partial template. The National Academy of Social Insurance provides a detailed roadmap for the development for financially sustainable state-based social insurance programs (Veghte et al. 2019).

### **Recognize, Redistribute, and Redesign**

Rather than treating care provision merely as a source of subjective satisfaction, we could recognize its moral valence and productive contributions. Rather than valuing human capital as an input into market output, we could value market output as an input into the improvement of human capabilities. Rather than fighting over the distribution of costs and benefits, we could emphasize the gains from well-designed social insurance programs. We cannot accomplish these goals without expanding the boundaries of production to look inside households and beyond market dynamics. Concerns about the negative consequences of inadequate

care provision in the United States are growing. As one example, a coalition of large corporations, small businesses, entrepreneurs, and investors formed a Care Economy Business Council in 2021 (<https://timesupnow.org/care-economy-business-council>) to “shift the cultural narrative about who is responsible for care, encourage and enforce equitable practices to support caregivers, and advocate for key public policy interventions.”

The research reviewed above reveals new ways of thinking about both private and public choices. More sustained attention to the economics of care provision could yield rich insights. Macroeconomic models could build on revised national accounting frameworks to consider social contracts, cultural norms, and consequences of pro-social preferences. In research as well as daily life, care pays forward.

■ *I gratefully acknowledge comments and suggestions from Franziska Dorn, Robert Frank, Shawn Fremstad, Leila Gautham, Misty Heggeness, Carol Heim, Anamary Maqueira Linares, Gregor Semieniuk, and Douglas Wolf, and deeply appreciate editorial feedback from Erik Hurst, Nina Pavcnik, Timothy Taylor, and Heidi Williams. This research enjoyed generous support from the Program on Gender and Care Work of the Political Economy Research Institute of the University of Massachusetts Amherst.*

## References

- Aguiar, Mark, and Erik Hurst.** 2005. “Consumption versus Expenditure.” *Journal of Political Economy* 113 (5): 919–48.
- Aguiar, Mark, Erik Hurst, and Loukas Karabarbounis.** 2013. “Time Use during the Great Recession.” *American Economic Review* 103 (5): 1664–96.
- Aizer, Anna, Hilary W. Hoynes, and Adriana Lleras-Muney.** 2022. “Children and the US Social Safety Net: Balancing Disincentives for Adults and Benefits for Children.” *Journal of Economic Perspectives* 36 (2): 149–74.
- Albelda, Randy.** 2011. “Time Binds: US Antipoverty Policies, Poverty, and the Well-Being of Single Mothers.” *Feminist Economics* 17 (4): 189–214.
- American Medical Association.** 2023. “Trends in Health Care Spending.” <https://www.ama-assn.org/about/research/trends-health-care-spending#:~:text=In%202021%2C%20spending%20was%20%24187.6,2021%2C%20spending%20was%20%2471.9%20billion> (accessed June 17, 2023).
- Arora, Kanika, and Douglas A. Wolf.** 2018. “Does Paid Family Leave Reduce Nursing Home Use? The California Experience.” *Journal of Policy Analysis and Management* 37 (1): 38–62.
- Associated Press.** 2002. “Feds Study Grandparents as Caregivers.” *New York Times*, June 4. <http://globalag.igc.org/elderrights/us/grandparentscaregivers.htm>.
- Badgett, M. V. Lee, and Nancy Folbre.** 2003. “Job Gendering: Occupational Choice and the Marriage Market.” *Industrial Relations* 42 (2): 270–98.
- Bailey, Martha J., Thomas E. Helgerman, and Bryan A. Stuart.** 2023. “How the 1963 Equal Pay Act and 1964 Civil Rights Act Shaped the Gender Gap in Pay.” NBER Working Paper 31332.

- Bailey, Martha J., Brad Hershbein, and Amalia R. Miller. 2012. "The Opt-In Revolution? Contraception and the Gender Gap in Wages." *American Economic Journal: Applied Economics* 4 (3): 225–54.
- Bartik, Timothy J. 2014. *From Preschool to Prosperity: The Economic Payoff to Early Childhood Education*. Kalamazoo, MI: Upjohn Institute.
- Basu, Rashmita, Adrienne C. Steiner, and Alan B. Stevens. 2022. "Long-Term Care Market Trend and Patterns of Caregiving in the US." *Journal of Aging & Social Policy* 34 (1): 20–37.
- Becker, Gary S., and Kevin M. Murphy. 1988. "The Family and the State." *Journal of Law and Economics* 31 (1): 1–18.
- Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income within Households." *Quarterly Journal of Economics* 130 (2): 571–614.
- Bittman, Michael, Paula England, Liana Sayer, Nancy Folbre, and George Matheson. 2003. "When Does Gender Trump Money? Bargaining and Time in Household Work." *American Journal of Sociology* 109 (1): 186–214.
- Blau, Francine D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55 (3): 789–865.
- Bowles, Samuel, Herbert Gintis, and Melissa Osborne. 2001. "The Determinants of Earnings: A Behavioral Approach." *Journal of Economic Literature* 39 (4): 1137–76.
- Bridgman, Benjamin, Andrew Craig, and Danit Kanal. 2022. "Accounting for Household Production in the National Accounts." *Survey of Current Business* 102 (2).
- Budig, Michelle J., Melissa J. Hodges, and Paula England. 2019. "Wages of Nurturant and Reproductive Care Workers: Individual and Job Characteristics, Occupational Closure, and Wage-Equalizing Institutions." *Social Problems* 66 (2): 294–319.
- Bullinger, Lindsey Rose. 2019. "The Effect of Paid Family Leave on Infant and Parental Health in the United States." *Journal of Health Economics* 66: 101–16.
- Bureau of Economic Analysis (BEA). 2019. "Glossary: National Income and Product Accounts." <https://www.bea.gov/resources/methodologies/nipa-handbook/pdf/glossary.pdf> (accessed June 17, 2023).
- Bursztnyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. "‘Acting Wife’: Marriage Market Incentives and Labor Market Investments." *American Economic Review* 107 (11): 3288–3319.
- Case, Anne, and Angus Deaton. 2020. *Deaths of Despair and the Future of Capitalism*. Princeton, NJ: Princeton University Press.
- Chari, A. V., and Elsa Valli. 2021. "The Effect of Subsidized Childcare on the Supply of Informal Care: Evidence from Public Kindergarten Provision in the US." *Journal of Health Economics* 77: 102458.
- Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–79.
- Children’s Defense Fund. 2021. *The State of America’s Children 2021*. Washington, DC: Children’s Defense Fund.
- Congressional Budget Office. 2023. *The Budget and Economic Outlook 2023 to 2033*. Washington, DC: Congressional Budget Office.
- Cortés, Patricia, and José Tessada. 2011. "Low-Skilled Immigration and the Labor Supply of Highly Skilled Women." *American Economic Journal: Applied Economics* 3 (3): 88–123.
- Currie, Janet. 2020. "Child Health as Human Capital." *Health Economics* 29 (4): 452–63.
- Doty, Pamela. 2017. "Private Pay Home Care: New Models of Access and Service Delivery." *Public Policy and Aging Report* 27 (3): 111–20.
- Doyle, Joseph J., Jr., and H. Elizabeth Peters. 2007. "The Market for Foster Care: An Empirical Study of the Impact of Foster Care Subsidies." *Review of the Economics of the Household* 5: 329–51.
- Duffy, Mignon, Randy Albelda, and Clare Hammonds. 2013. "Counting Care Work: The Empirical and Policy Applications of Care Theory." *Social Problems* 60 (2): 145–67.
- England, Paula, Michelle Budig, and Nancy Folbre. 2002. "Wages of Virtue: The Relative Pay of Care Work." *Social Problems* 49 (4): 455–73.
- Folbre, Nancy. 2006. "Measuring Care: Gender, Empowerment, and the Care Economy." *Journal of Human Development* 7 (2): 183–99.
- Folbre, Nancy. 2009. *Greed, Lust, and Gender: A History of Economic Ideas*. New York: Oxford University Press.
- Folbre, Nancy, ed. 2012. *For Love and Money: Care Provision in the United States*. New York: Russell Sage Foundation.

- Folbre, Nancy.** 2021. *The Rise and Decline of Patriarchal Systems: An Intersectional Political Economy*. New York: Verso Books.
- Folbre, Nancy, Shawn Fremstad, and Pilar Gonalons-Pons.** 2023. "Measuring Care: Resources, Shortfalls, and Possible Improvements." University of Massachusetts Political Economy Research Institute Working Paper. <https://peri.umass.edu/publication/item/1735-measuring-care-provision-in-the-united-states-resources-shortfalls-and-possible-improvement>.
- Folbre, Nancy, Leila Gautham, and Kristin Smith.** 2022. "Gender Inequality, Bargaining, and Pay in Care Services in the United States." *ILR Review* 76 (1): 86–111.
- Folbre, Nancy, Janet C. Gornick, Helen Connolly, and Teresa Munzi.** 2013. "Women's Employment, Unpaid Work, and Economic Inequality." In *Income Inequality: Economic Disparities and the Middle Class in Affluent Countries*, edited by Janet Gornick and Markus Jännti, 234–60. Stanford, CA: Stanford University Press.
- Folbre, Nancy, Marta Murray-Close, and Jooyeoun Suh.** 2018. "Equivalence Scales for Extended Income in the U.S." *Review of Economics of the Household* 16 (2): 189–227.
- Folbre, Nancy, and Julie A. Nelson.** 2000. "For Love or Money—Or Both?" *Journal of Economic Perspectives* 14 (4): 123–40.
- Folbre, Nancy, Cordelia Reimers, and Jayoung Yoon.** 2009. "Making Do and Getting By: Non-market Work and Elderly Women's Standards of Living in the United States." *Journal of Women, Politics and Policy* 30 (2–3): 198–221.
- Folbre, Nancy, and Barnet Wagman.** 1993. "Counting Housework: New Estimates of Real Product in the United States, 1800–1860." *Journal of Economic History* 53 (2): 275–88.
- Folbre, Nancy, and Thomas Weisskopf.** 1998. "Did Father Know Best? Families, Markets, and the Supply of Caring Labor." In *Economics, Values and Organization*, edited by Avner Ben-Ner and Louis Putterman, 171–205. New York: Cambridge University Press.
- Folbre, Nancy, and Douglas Wolf.** 2013. "The Intergenerational Welfare State." *Population and Development Review* 38: 36–51.
- Fortin, Nicole M.** 2008. "The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People." *Journal of Human Resources* 43 (4): 884–918.
- Frazis, Harley, and Jay Stewart.** 2011. "How Does Household Production Affect Measured Income Inequality?" *Journal of Population Economics* 24 (1): 3–22.
- Fuchs, Victor R.** 1988. *Women's Quest for Economic Equality*. Cambridge, MA: Harvard University Press.
- Gál, Róbert Iván, Pieter Vanhuysse, and Lili Vargha.** 2018. "Pro-elderly Welfare States within Child-Oriented Societies." *Journal of European Public Policy* 25 (6): 944–58.
- Gautham, Leila, and Nancy Folbre.** Forthcoming. "Child Care, Household Expenditures and Consumption of Unpaid Services: Evidence from the U.S. Panel Study of Income Dynamics." *Review of Income and Wealth*.
- Goldin, Claudia.** 1990. *Understanding the Gender Gap: An Economic History of American Women*. Oxford, UK: Oxford University Press.
- Goldin, Claudia.** 2021. *Career and Family: Women's Century-Long Journey toward Equity*. Princeton, NJ: Princeton University Press.
- Goodman, Julia M., Dawn M. Richardson, and William H. Dow.** 2022. "Racial and Ethnic Inequities in Paid Family and Medical Leave: United States, 2011 and 2017–2018." *American Journal of Public Health* 112 (7): 1050–58.
- Grall, Timothy.** 2020. *Custodial Mothers and Fathers and Their Child Support: 2017*. Washington, DC: US Census Bureau.
- Hirsch, Barry T., and Julia Manzella.** 2015. "Who Cares—And Does It Matter? Measuring Wage Penalties for Caring Work." In *Gender Convergence in the Labor Market*, edited by Solomon W. Polachek, Konstantinos Tatsiramos, and Klaus F. Zimmermann, 213–75. Bingley, UK: Emerald.
- Huang, Sean Shenghsiu, and John R. Bowblis.** 2019. "Is the Quality of Nursing Homes Countercyclical? Evidence from 2001 through 2015." *Gerontologist* 59 (6): 1044–54.
- Hudson, Peter, R. Sean Morrison, Richard Schulz, Abraham Aizer Brody, Constance Dahlin, Kathleen Kelly, and Diane E. Meier.** 2020. "Improving Support for Family Caregivers of People with a Serious Illness in the United States: Strategic Agenda and Call to Action." *Palliative Medicine Reports* 1 (1): 6–17.
- Ishizuka, Patrick.** 2021. "The Motherhood Penalty in Context: Assessing Discrimination in a Polarized Labor Market." *Demography* 58 (4): 1275–1300.
- Johnson, Richard W., Karen E. Smith, and Barbara A. Butrica.** 2023. *Lifetime Employment-Related Costs to*

- Women of Providing Family Care*. Washington, DC: Urban Institute.
- Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler.** 1990. "Experimental Tests of the Endowment Effect and the Coase 5 Theorem." *Journal of Political Economy* 98 (6): 1325–48.
- Kamas, Linda, and Anne Preston.** 2020. "Does Empathy Pay? Evidence on Empathy and Salaries of Recent College Graduates." *Journal of Labor Research* 41 (1–2): 169–88.
- Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller.** 2019. "Child Penalties across Countries: Evidence and Explanations." *American Economic Review Papers and Proceedings* 109: 122–26.
- Konish, Lorie.** 2022. "Americans Are Eager for More Monthly Child Tax Credit Checks." *CNBC*, March 1. <https://www.cnbc.com/2022/03/01/future-child-tax-credit-payments-could-come-with-work-requirements.html>.
- Lambert, Susan J., Julia R. Henly, and Jaeseung Kim.** 2019. "Precarious Work Schedules as a Source of Economic Insecurity and Institutional Distrust." *RSF: Russell Sage Foundation Journal of the Social Sciences* 5 (4): 218–57.
- Landefeld, J. Steven, Barbara M. Fraumeni, and Cindy M. Vojtech.** 2009. "Accounting for Household Production: A Prototype Satellite Account Using the American Time Use Survey." *Review of Income and Wealth* 55 (2): 205–25.
- Landefeld, J. Steven, and Stephanie H. McCulla.** 2000. "Accounting for Nonmarket Household Production within a National Accounts Framework." *Review of Income and Wealth* 46 (3): 289–307.
- Lee, Ronald, David McCarthy, James Sefton, and Jože Sambt.** 2017. "Full Generational Accounts: What Do We Give to the Next Generation?" *Population and Development Review* 43 (4): 695–720.
- Lino, Mark, Kevin Kuczynski, Nestor Rodriguez, and TusaRebecca Schap.** 2017. *Expenditures on Children by Families, 2015*. Washington, DC: U.S. Department of Agriculture, Center for Nutrition Policy and Promotion.
- Meyer, Bruce D., and James X. Sullivan.** 2008. "Changes in the Consumption, Income, and Well-Being of Single Mother Headed Families." *American Economic Review* 98 (5): 2221–41.
- Miller, Grant.** 2008. "Women's Suffrage, Political Responsiveness, and Child Survival in American History." *Quarterly Journal of Economics* 123 (3): 1287–1327.
- Munnell, Alicia H., and Andrew D. Eschtruth.** 2018. "Modernizing Social Security: Caregiver Credits." Center for Retirement Research at Boston College Issue in Brief 18–15.
- Myers, Caitlin Knowles, and Morgan Welch.** 2021. "What Can Economic Research Tell Us about the Effect of Abortion Access on Women's Lives." Brookings Institution, November 30. <https://www.brookings.edu/research/what-can-economic-research-tell-us-about-the-effect-of-abortion-access-on-womens-lives/>.
- National Academies of Sciences, Engineering, and Medicine.** 2019. *A Roadmap to Reducing Child Poverty*. Washington, DC: National Academies Press.
- Nussbaum, Martha C.** 2011. *Creating Capabilities*. Cambridge, MA: Harvard University Press.
- Petts, Richard J., Chris Knoester, and Qi Li.** 2020. "Paid Paternity Leave-Taking in the United States." *Community, Work and Family* 23 (2): 162–83.
- Reinhard, Susan C., Lynn Friss Feinberg, Ari Houser, and Molly Evans.** 2019. "Valuing the Invaluable 2019 Update: Charting a Path Forward." Washington, DC: AARP Public Policy Institute.
- Rodakowski, Juleen, Philip B. Rocco, Maqui Ortiz, Barbara Folb, Richard Schulz, Sally C. Morton, Sally Caine Leathers, Lu Hu, and A. Everette James III.** 2017. "Caregiver Integration During Discharge Planning for Older Adults to Reduce Resource Use: A Metaanalysis." *Journal of the American Geriatrics Society* 65 (8): 1748–55.
- Romero, Mary.** 2018. "Reflections on Globalized Care Chains and Migrant Women Workers." *Critical Sociology* 44 (7–8): 1179–89.
- Ruppanner, Leah, Caitlyn Collins, Liana Christin Landivar, and William J. Scarborough.** 2021. "How Do Gender Norms and Childcare Costs Affect Maternal Employment across US States?" *Gender and Society* 35 (6): 910–39.
- Sen, Amartya.** 1999. *Development as Freedom*. New York: Alfred A. Knopf.
- Stiglitz, Joseph E., Jean-Paul Fitoussi, and Martine Durand.** 2019. *Measuring What Counts: The Global Movement for Well-Being*. New York: New Press.
- Stone, Deborah.** 2000. "Caring by the Book." In *Care Work: Gender, Labor and the Welfare State*, edited by Madonna Harrington Meyer, 89–111. New York: Routledge.
- Suh, Jooyeoun, and Nancy Folbre.** 2016. "Valuing Unpaid Child Care in the U.S.: A Prototype Satellite Account Using the American Time Use Survey." *Review of Income and Wealth* 62 (4): 668–84.

- Suh, Jooyeoun, and Nancy Folbre.** 2023. “The Responsibilities of Parental Childcare: Evidence from the American Time Use Survey.” Paper presented at the annual meeting of the Allied Social Science Associations, New Orleans, January 8.
- Urban Institute.** 2023. “State and Local Backgrounders.” <https://www.urban.org/policy-centers/cross-center-initiatives/state-and-local-finance-initiative/state-and-local-backgrounders/higher-education-expenditures#Question1Higher> (accessed June 17, 2023).
- US Bureau of Labor Statistics.** 2022a. “American Time Use Survey—2021 Results.” June 23. <https://www.bls.gov/news.release/atus.toc.htm>.
- US Bureau of Labor Statistics.** 2022b. “Labor Force Statistics from the Current Population Survey.” <https://www.bls.gov/cps/cpsaat18.htm> (accessed June 30, 2022).
- US Supreme Court.** 1973. *Row v. Wade*.
- US Supreme Court.** 2022. *Dobbs v. Jackson Women’s Health Organization*.
- US Treasury.** 2021. “The Economics of Child Care Supply in the United States” <https://home.treasury.gov/system/files/136/The-Economics-of-Childcare-Supply-09-14-final.pdf> (accessed July 18, 2023).
- Veghte, Benjamin W., Alexandra L. Bradley, Marc Cohen, Heidi Hartmann, eds.** 2019. *Designing Universal Family Care: State-Based Social Insurance Programs for Early Child Care and Education, Paid Family and Medical Leave, and Long-Term Services and Supports*. Washington, DC: National Academy of Social Insurance.
- Vespa, Jonathan, Lauren Medina, and David M. Armstrong.** 2018. *Demographic Turning Points for the United States: Population Projections for 2020 to 2060*. Washington, DC: US Department of Commerce, Economics and Statistics Administration, US Census Bureau.
- Wagman, Barnet, and Nancy Folbre.** 1996. “Household Services and Economic Growth in the United States, 1870–1930.” *Feminist Economics* 2 (1): 43–66.
- Wilkinson, Richard, and Kate Pickett.** 2011. *The Spirit Level: Why Greater Equality Makes Societies Stronger*. New York: Bloomsbury Press.
- Wolf, Douglas, and Nancy Folbre, eds.** 2012. *Universal Coverage of Long-Term Care in the United States: Can We Get There from Here*. New York: Russell Sage.
- Wolf, Douglas A., Ronald D. Lee, Timothy Miller, Gretchen Donehower, and Alexandre Genest.** 2011. “Fiscal Externalities of Becoming a Parent.” *Population and Development Review* 37 (2): 241–66.

# Job Training and Job Search Assistance Policies in Developing Countries

Eliana Carranza and David McKenzie

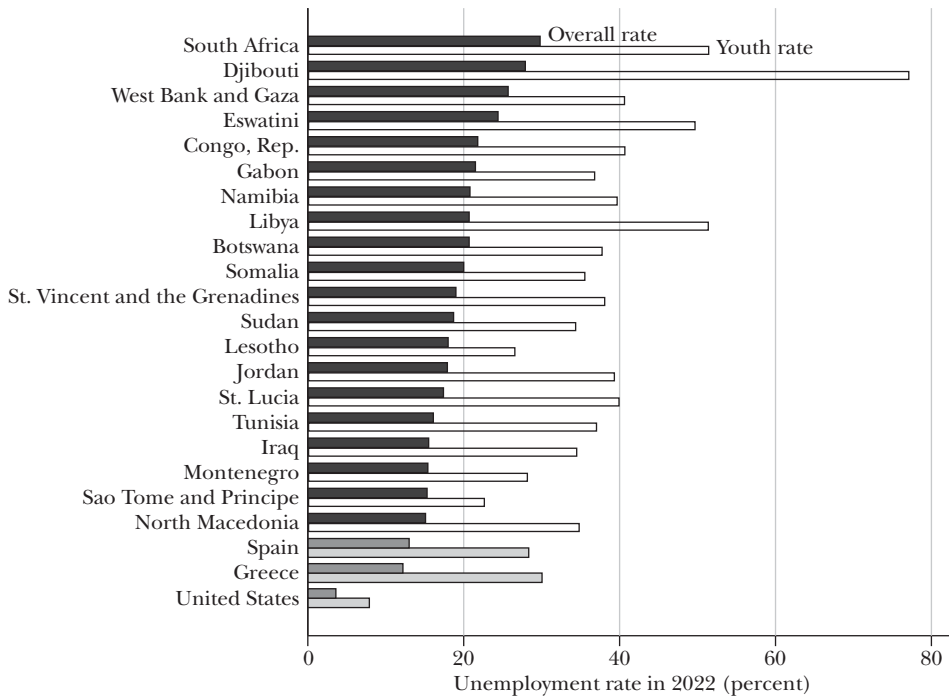
**I**n countries across the developing world, headlines commonly warn of a combined jobs crisis and demographic time bomb, in which millions of jobs need to be created each year. For example, “India is sitting on a time bomb’: Jobs crisis looms as population soars” warns “The country needs to create at least 90 million new non-farm jobs by 2030 to absorb new workers” (NewsIn. Asia 2023). “Africa’s Youth Unemployment Crisis Is a Global Problem” notes that “while 10 million to 12 million youth enter the workforce in Africa each year, only 3 million formal jobs are created annually” (Donkor 2021). Along with the issues posed by new job seekers in the future, unemployment of existing workers is already high in some countries. Despite a perception that unemployment rates tend to be low in developing countries because people can always work in agriculture or are too poor not to work, the 20 countries with the highest unemployment rates in the world in 2022 were all in the developing world, as shown in Figure 1. Unemployment rates exceeded 15 percent, higher than the 13 percent in Spain and 12 percent in Greece, the European Union’s two highest rates, and far in excess of the 3.6 percent unemployment rate in the United States. They were even higher for youth, exceeding 30 percent in most of these 20 countries.

■ *Eliana Carranza is a Global Lead for Labor Markets and Skills in the Social Protection and Jobs Unit, and David McKenzie is a Lead Economist in the Development Research Group, both at the World Bank, Washington, DC. Their email addresses are [ecarranza@worldbank.org](mailto:ecarranza@worldbank.org) and [dmckenzie@worldbank.org](mailto:dmckenzie@worldbank.org).*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.1.221>.

Figure 1

### The Countries with the Highest Unemployment Rates Are All in the Developing World



Source: World Bank (2023).

Note: Unemployment refers to the share of the total labor force that is without work but available for and seeking employment. Youth Unemployment is similarly defined for the share of the total labor force ages 15–24.

Even among those with employment, many find themselves in low paid, informal jobs, and would like to do better. Changes in the structure of the economy and job opportunities due to automation, climate change, and lasting impacts of the COVID-19 pandemic are thought to have made the challenge even harder, resulting in headlines such as “Robots Pose Big Threat to Jobs in Africa, Researchers Warn” (Ridgwell 2018), “Rising heat stress could destroy 80 million jobs by 2030, UN says” (Taylor 2019), and “This Chinese jobs crisis could be its worst” (Chen 2022).

Faced with these current employment challenges and such forecasts, governments face pressure to help jobseekers. One viewpoint is that the problem is a skills mismatch, where many of those seeking jobs do not have the skills sought by employers. This problem may be due to poor education systems as well as changes in skills demanded by employers as economic growth, technological change, globalization, and the desire for a green transition change the structure of the economy. A



popular solution is then for the government to provide *job training* so that jobseekers can acquire these skills. An alternative viewpoint is that even when workers have skills that employers want, they have difficulty finding the right job fit due to search and matching frictions. Fragmented and largely informal labor markets may make it difficult to identify job openings, and workers without much work experience may not know how best to look for jobs or signal their skills to potential employers. This results in governments providing a range of *job search assistance* policies designed to help jobseekers better use the skills they do have.

This paper asks whether, when, and how developing country governments should undertake job training and job search assistance policies. Job generation will typically require policies that boost the demand for labor by increasing firm productivity and overall economic growth. Supply-side interventions that train workers and help them look for jobs will not be very effective if there are few jobs for them. An earlier generation of critical reviews and meta-analyses of the first generation of evaluations of these programs found the typical impacts were rather small, with typically only two or three people out of every 100 trained or assisted finding work as a result of these programs (McKenzie 2017). As a result, Blattman and Ralston (2015) argued that “it is hard to find a skills training program that passes a simple cost-benefit test.”

However, there have been recent innovations in how job training and job search programs in developing countries are designed, targeted, and implemented.<sup>1</sup> We argue that governments can play a real, if more limited, role in providing these programs. For example, government action may be warranted when private firms are underinvesting in training of workers due to the possibility workers will leave, or when search and matching frictions slow down or prevent the reallocation of workers across sectors and geography that are critical for structural transformation. Equity concerns may also justify the use of these policies to help disadvantaged jobseekers find jobs. We begin with an overview of the worse and better reasons why governments in developing countries have become involved in job search and job training. We then outline lessons from recent experiences in developing economies with vocational training, policies designed to overcome spatial and informational search frictions, interventions aimed at overcoming psychological barriers to job search, and government efforts to encourage the development and use of online jobs platforms.

<sup>1</sup>We focus on job training and job search assistance as two of the most common types of active labor market policies undertaken by governments. We do not cover several other types of active labor market policy. For example, another common form of skills training is to teach business skills to jobseekers in order that they can start their own small businesses (for a review, see McKenzie et al. 2021). Governments can also directly provide jobs to workers through large public works programs (reviewed in Gehrke and Hartwig 2018) or make it cheaper for firms to hire workers through wage subsidy programs (reviewed in McKenzie 2017). We also restrict our focus to largely urban labor markets, reflecting government efforts to integrate workers into wage jobs, and do not discuss policies to increase agricultural employment.

Table 1

**Main Methods Used by Employed Workers to Find Jobs**

	Percent who used method to find job						
	Albania	Jordan	Mexico	Morocco	Romania	Sierra Leone	Turkey
Friends and relatives	28.8	41.0	53.0	70.9	37.2	46.0	59.7
Direct application to employers	52.7	37.1	32.0	40.4	36.4	41.1	22.9
Newspaper ads	0.5	8.5	8.9	0.1	15.7	5.0	4.5
Internet	0.3	1.0	1.6	1.5		0.0	5.0
Public or private employment agency	9.1	6.2	0.6	5.5	2.5	2.2	2.7
Other	8.5	6.2	3.9	12.5	8.2	5.8	5.2
Sample size	14,653	904	117,924	712	36,542	504	3,302

*Source:* Albanian data from Quarterly Labor Force Survey 2019 (ILOSTAT 2019); Jordan is from New Work Opportunities for Women Pilot Impact Evaluation 2010–2013 (Groh et al. 2013); Mexico is from Trimester 1, 2014 National Survey of Occupation and Employment (ENOE) (INEGI 2023); Morocco is from Household and Youth Survey 2009–2010 (World Bank 2013); Romania is from Household Labor Force Survey 2021 (ILOSTAT 2021); Sierra Leone is from 2014 Labor Force Survey (Statistics Sierra Leone 2014); and Turkey is from Vocational Training for the Unemployed Impact Evaluation 2010–2012 (McKenzie et al. 2014).

*Note:* Morocco survey allowed multiple methods to be used, so responses add up to more than 100 percent. Romania survey combines newspaper advertisements and internet advertisements into one response category. Carranza and McKenzie (2023) provide full details.

## Good and Not-So-Good Reasons for Governments to Provide Job Training and Search Assistance

Firms typically have strong incentives to find and retain good workers, and many workers likewise have strong incentives to look for jobs and seek training in the skills demanded by the market. As a result, the vast majority of jobs are filled without government intervention. Table 1 illustrates this point, using some of the relatively few available labor force surveys that directly ask employed workers in developing countries how they found their jobs. The vast majority of jobs are found by workers either learning about jobs through their social networks of friends and relatives, or through workers directly approaching employers in their business or worksite and asking if they have openings. Public (government) agencies and private employment agencies are the source of a tiny fraction of all jobs found by workers: 9 percent of jobs in Albania, 5 to 6 percent of jobs in Jordan and Morocco, and down to less than 1 percent of jobs obtained in Mexico.

Job matches between workers and firms can be highly location-, sector-, time-, and firm-specific, so that it seems unlikely in most cases that a central planner or government will do better than the private market in filling most job openings. In line with Table 1, throughout history and the growth of most countries, large increases in labor supply have been absorbed by the market without the need for the government to help millions of people find jobs. So why should the government get involved in providing job training and search assistance rather than leaving it to the private market?

Perhaps the most common reason that governments provide these services is based on political pressure for a government to show it is doing something about employment—and providing job training or search assistance is often politically easier than addressing other barriers to creating jobs. In political terms, the success of these programs can be based on visible inputs (like the number of people who finished a training program, or the number of towns with job-assistance centers) rather than the more difficult-to-observe effects like the number of workers who obtain lasting good jobs that they would not have found without this help. When quantitative inputs are the measure of success, the consequence can be reliance on public sector training agencies that may have limited linkages to the private sector and that offer training of variable quality that is not necessarily in skills that are in high demand. For example, Maitra, Maitra, and Thakur (2022) note that, despite efforts at reform, India’s system of vocational training is still characterized by “a bureaucracy-driven centralised and hierarchical framework” with a “glaring lack of involvement of industries,” resulting in a system that “has not proved agile enough to quickly adapt skill-training provisions to contemporary technological innovations.”

An additional not-so-good reason for providing these job training and search services is as a second-best response to distortions in the labor market created by other government actions that may be politically much more difficult to change. One example is the dominance of a large public sector that pays much higher wages than comparable private sector jobs, causing jobseekers to queue for these public jobs and not consider working in the private sector. For example, this pattern has historically defined the job market for relatively educated workers in many countries in the Middle East. Assaad (2014) notes that “by using labor markets as means to distribute rents and to buy political quiescence, Arab governments have essentially undermined the labor markets’ primary function.” Another example is the presence of high minimum wages and inflexible labor laws that make it expensive and burdensome for private employers to hire workers, leading the supply of jobseekers to greatly exceed demand at prevailing wages. South Africa introduced a minimum wage set roughly equal to the median wage, and soon had an unemployment rate over 30 percent (Bhorat, Lilenstein, and Stanwix 2021).<sup>2</sup> Many of these regulations are size-dependent, applying more strictly to firms that have more than a given threshold (say, 10, 50, or 100) number of workers, acting as an additional constraint on the expansion of more productive, larger firms.<sup>3</sup> While

<sup>2</sup>Of course, quantifying the impact of the minimum wage on employment is difficult. Bhorat et al. (2021) suggest that high levels of noncompliance with the policy may mute some of the effects, although their analysis may be too soon for the full effects to be felt. Bertrand and Crépon (2021) provide experimental evidence that South African employers imperfectly understand labor regulations and this constrains their hiring.

<sup>3</sup>Firms often find workarounds, so these regulations may bind less strongly than in developed countries. For example, Bertrand, Hsieh, and Tsivanidis (2022) detail how Indian firms use outside contract labor to overcome some of the burdensome labor regulations that otherwise would apply when they reach 100 workers.

training and search assistance may help some workers to overcome the constraints caused by these distortions and find jobs, any success stories may just crowd out other jobseekers, and it would be better for governments to concentrate their policy efforts in addressing the distortions directly.

In the absence of these government-introduced distortions, the simplest introductory model of labor supply and labor demand, in which the market works to equilibrate the supply and demand for jobs through changes in the wage, is actually not a bad approximation for many types of labor in less-skilled jobs in urban labor markets in developing countries. Many firms have no trouble filling the job openings that they have. For example, Groh et al. (2015) conducted a panel survey of employers in Jordan to track how long it took firms to fill vacancies. Most vacancies are filled fast, with only 6 percent requiring more than two weeks. In a field experiment conducted in Sri Lanka, de Mel, McKenzie, and Woodruff (2019) find that only 12 percent of microenterprises said they found it hard to find the right worker for routine, physical jobs. Many governments are particularly interested in boosting the workforce in formal jobs in large manufacturing firms. Several studies, however, have found these manufacturing firms do not appear to be that discerning about who they hire. They find workers very quickly and many workers voluntarily quit these jobs—suggesting that these jobs are neither that hard to fill, nor so rationed and sought after that no one leaves these jobs after securing such a position (for example, Blattman and Dercon 2018).<sup>4</sup>

So then what are some better reasons for government involvement in the provision of job training and search assistance? There are three main reasons why the private market allocation of labor may be inefficient or inequitable, thus providing a potential justification for the government to get involved. First, search and matching frictions are likely to exist for some jobs, and they not only inhibit individual jobseekers from finding work, but can slow down or prevent the reallocation of workers across sectors and geography that is such an integral part of structural transformation. For example, China's rapid growth involved its urban population growing from 100 million in 1980 to over 500 million today. This enormous reallocation involves workers needing to learn about job opportunities in another place, and potentially also requires them to learn different skills. The government may be able to speed up this transition process through appropriate training and job search policies. Matching frictions can also mean inefficiently high rates of job turnover, which can be costly for both workers and firms, and so efforts to improve match quality may increase labor productivity.

Second, the production of human capital can involve externalities that cause firms and individual workers to underinvest in training. For example, looking at empirical evidence from Colombia, Caicedo, Espinosa, and Seibold (2022) argue that firms underinvest in the training of workers due to the possibility that they will

<sup>4</sup>Of course, this high turnover can have costs for both firms and workers, and better matches can result in higher productivity. But many firms do not appear to view these costs as high enough to invest heavily in reducing turnover.

leave and work for other firms. Likewise, the apprenticeship system common in West Africa may embed inefficiencies caused by the concern of master craftspeople that their apprentices will compete against them after they have been trained—which creates incentives for firm owners to provide slower and lower-quality training to apprentices than they would do if properly incentivized (Brown et al. forthcoming). Government support of training in general, and for sector-specific (rather than firm-specific) skills, may address these concerns. Concerted investment in certain sets of skills may also be needed as part of an industrial policy to attract large multinational plants that have spillover benefits for the rest of the economy. For example, the Costa Rican Investment Promotion Agency (CINDE) has worked with partners to provide training in technological fields demanded by multinational companies.

Finally, governments may wish to be involved for social mobility and equity reasons. As shown in Table 1, the main way many firms and jobseekers connect is through networks of connections formed by friends and relatives. But disadvantaged individuals with limited networks and a lack of funds to spend acquiring skills may end up segmented into different labor markets, without the knowledge or skills to approach firms directly and find jobs that way. Even if training and job search assistance to these groups does not generate additional jobs, providing an opportunity for some individuals from these disadvantaged groups to access better jobs may be desirable for equity reasons, and may help improve overall allocative efficiency in the economy.

While our focus is on low- and middle-income economies, the developing world displays a wide range of labor market conditions, and many of these same issues and rationales for government intervention apply to certain labor market segments in high-income countries. Similar concerns apply about whether workers displaced by trade or technology shocks, or stuck in cities with declining industries, are able to reskill and relocate, as well as equity reasons being used to focus programs on individuals from disadvantaged backgrounds. However, a larger share of the labor market may be subject to these frictions in many developing countries, which in addition to greater demographic pressures and an ongoing structural transformation of the economy can strengthen the rationale for policy action.

Given these reasons that might justify government involvement, the key question is how should the government get involved in order to ensure these policies have higher chances of being successful?

## **What Are the Most Promising Avenues for Government Involvement in Job Training and Job Search?**

Jobseekers may struggle to find jobs for three main reasons. First, there may be a shortage of firms wishing to hire them. This may reflect a lack of overall labor demand in the economy, in which case the appropriate policy actions will be in the area of private sector development policy. In some cases, it may also reflect discrimination, with firms not willing to hire individuals with certain characteristics. Second,

jobseekers may struggle because they lack the skills and experience needed for jobs. Job training and internships and apprenticeships can then be used to help overcome this problem. Finally, even if jobs are available and individuals have the right skills, they may struggle to find and match with employers who want their labor, in which case job search assistance can be useful.

It is difficult to find systematic evidence to assess the relative importance of these three reasons, and the answer will almost surely vary with context. Our sense is that in many cases the largest issue is a lack of labor demand for jobs, especially “good” jobs, and that policies to help firms to grow and demand more labor need to be a primary part of any jobs policy solution (although it is not the topic of this paper). Data from the Mexican National Survey of Occupation and Employment (ENOE) provides one data point, asking unemployed individuals why they are not looking for work: 4.6 times as many individuals say it is because they think there are no jobs available near them than say it is because they lack the skills or experience needed for jobs. Lack of experience and knowledge of where to look for jobs may be bigger barriers for young jobseekers. A survey of high school graduates in one part of Mexico with a strong labor market found 33 percent of youth said lack of experience was their main obstacle to finding a job, 10 percent said lack of skills, and 14 percent said difficulty searching for jobs.<sup>5</sup>

While employers can often fill jobs with “a” worker quickly, many employers say that difficulty finding workers with the right skills is an issue they face, and many employers appear to feel this is more the case today than in the past. Manpower Group has been surveying public and private employers in multiple countries annually and reports that 77 percent of employers in 41 countries surveyed in 2023 report difficulty finding talent with the skills they need, up from only 31 percent in 2010. Figure 2 shows high levels of firms saying they struggle to find talent in all of the developing countries the survey covers, with the percentage of firms reporting difficulties similar to that in the United States and Germany. Employers report both difficulty finding workers with the right technical skills, such as information technology, data skills, and sales and marketing skills, as well as with the right soft skills, such as reliability, creativity, critical thinking, and resilience. Public policy efforts can then try to help jobseekers develop these skills, as well as helping workers with these skills to signal in a credible way that they have these skills as they search for jobs and match with employers.

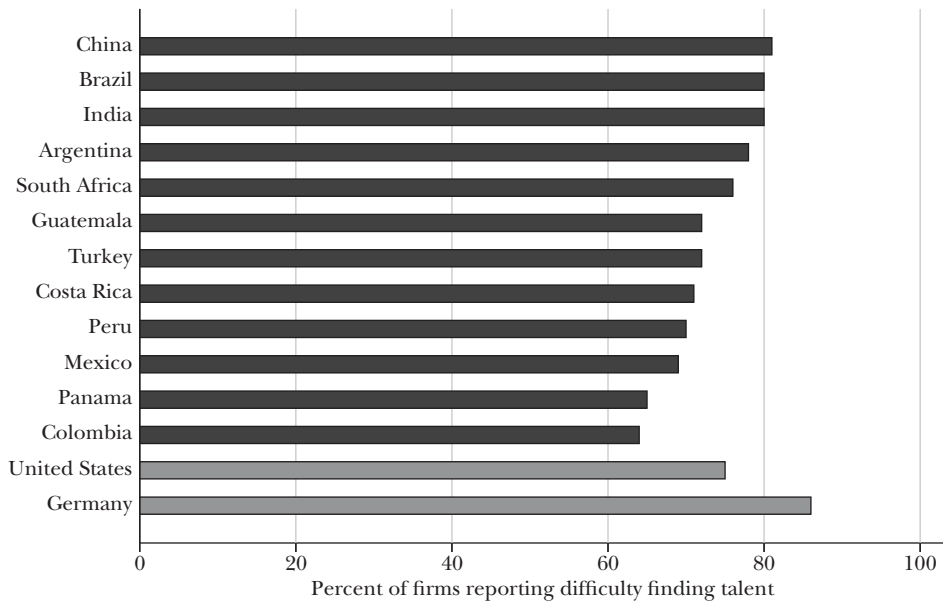
### **Job Training**

Job training programs are designed to provide new skills and experience, and are predominantly focused on youth and the unemployed. They typically try to teach technical vocational skills in fields such as hairdressing and beauty, carpentry,

<sup>5</sup>Data from México Piloto de Inclusion Laboral de Jóvenes Baseline survey Phase 1 (2018), used in Abel et al. (2022). A further 21 percent said their age, gender, ethnicity, or socioeconomic background was the main obstacle, which could reflect employer discrimination, social norms, or difficulties they face paying for training or job search.

*Figure 2*

**The Majority of Firms in Many Developing Countries Say They Struggle to Find Talent with the Skills They Need**



*Source:* Manpower Group (2023).

*Note:* Percentages are calculated as the share of employers in a representative sample of public and private employers for the country, across specific sectors and regions, reporting a labor demand that outpaces the supply of labor with the skills they need at their location during the second quarter of 2023.

electrical work, tailoring, plumbing, or information technology skills such as coding, data entry, office programs, and others. In addition to these “hard” skills, some programs also include “soft skills” components such as communication skills, teamwork, planning, self-efficacy, and financial literacy. These programs can be taught in classrooms and/or in the form of on-the-job-training through an internship or apprenticeship. The most common programs offered by governments and studied in the literature tend to last three to six months, although there are shorter intensive “boot camps” that can be three weeks to a month, as well as longer programs of two years that are more similar to the traditional Technical and Vocational Education and Training (TVET) programs that are sometimes offered as part of the formal education system.

The hope is that these programs will increase employment and earnings for jobseekers through at least three potential channels: (1) by increasing their human capital through teaching new skills, thereby making them more productive workers; (2) by alleviating employer uncertainty about the skills workers have by providing a signal in the form of certification or references; (3) by providing jobseekers with

new strategies (and potentially new networks) for finding jobs. Public funding of these programs is often justified by arguing that jobseekers are credit-constrained and unable to pay for the direct costs of training, as well as unable to bear the opportunity costs involved in needing to pay for living expenses and not earning money during training. In addition, due to information frictions, jobseekers may not know about the full range of training providers or find it difficult to ascertain their quality. These same arguments are also made in high-income countries, but credit constraints and informational frictions are likely to be larger issues in less developed economies. On the firm side, even though firms may have trouble finding workers with the right skills, firms also may be credit-constrained in paying for worker training and reluctant to spend time and money training workers in general skills if these workers may then leave to work in other firms soon afterwards.

Returns to education and experience are among the strongest empirical regularities in labor economics, suggesting that training should affect earnings and employment. However, given that these courses are short in duration and returns to education and experience are typically in the order of around 10 percent per year, we might expect only three or six months of training to have relatively modest impacts. Indeed, most randomized experiments measuring the impact of vocational training programs have found effects of roughly this size. McKenzie (2017) reviews nine such studies and finds an average impact of a 2.3 percentage point increase in employment, which, given the costs of these programs, equates to approximately \$17,000–\$60,000 per additional person employed, and a median increase in earnings of 11 percent, or \$19 per month. Agarwal and Mani (2023) include an additional 14 recent studies in a formal meta-analysis, and find an average impact on employment of 4 percentage points (with a 95 percent confidence interval between 2 to 6 percentage points) and on earnings of 8.2 percent (with a 95 percent confidence interval between 2 to 14 percent). These impacts on employment are similar in magnitude to the average impact of training in high-income countries, with Card, Kluve, and Weber (2018) finding from a combination of experimental and nonexperimental studies that training averages a 2.0 percentage point impact on employment in the first year after training, and 6.6 percentage points one to two years later.<sup>6</sup>

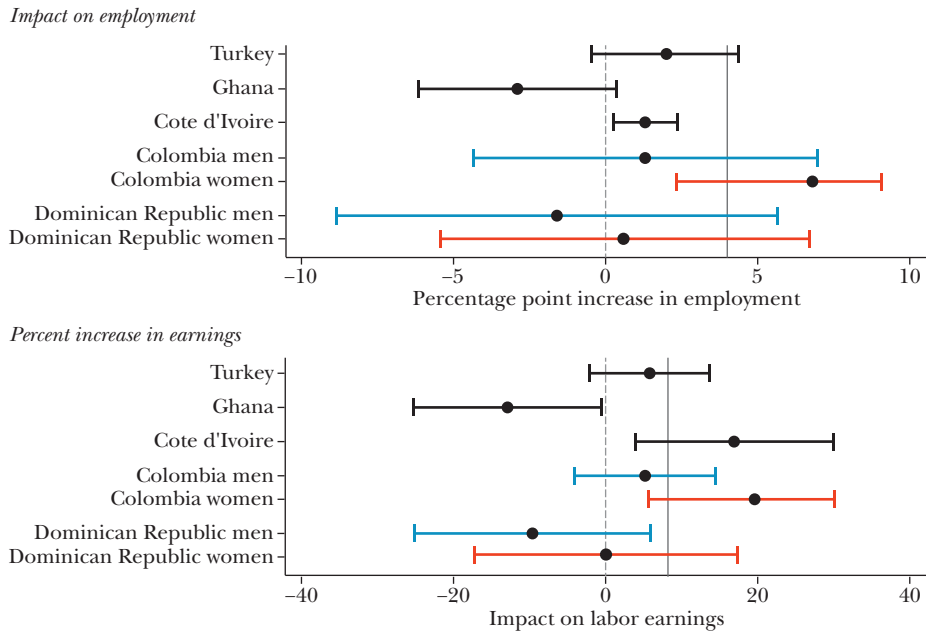
These effects appear particularly muted for at-scale programs operated by governments. Figure 3 shows evidence from five experimental evaluations of government vocational training programs that trained at least 5,000 people in a year. In this figure and the subsequent ones, we show both the point estimate and a 95 percent confidence interval from the evaluation. For example, the figure shows that vocational training in Turkey resulted in an estimated 2.0 percentage point increase in employment, with a 95 percent confidence interval of –0.5 to

<sup>6</sup>Card, Kluve, and Weber (2018) use a sample that combines developed and developing countries, with 64 percent of the training estimates coming from high-income. They find few significant differences in impact by country region, except for slightly higher estimates from nonexperimental evaluations in Germanic countries.



Figure 3

**Limited Effects of Large-Scale Government Job Training Programs on Employment and Earnings**



Source: Turkey estimates from Hirschleifer et al. (2016) at one year post-training; Ghana estimates from Hardy et al. (2019) at one year post-training; Côte d'Ivoire estimates from Crépon and Premand (2019) at two years post-training; Colombia estimates from Attanasio, Kugler, and Meghir (2011) at 14 months post-training; Dominican Republic estimates from Acevedo et al. (2020) at 3.5 years post-training. Solid vertical lines show meta-analysis estimates from Appendix Figures A.1 and A.2 in Agarwal and Mani (2023) from a broader set of government and non-government programs, including pilot programs: of 4.0 percentage points for employment and 8.2 percent for earnings.

Notes: Studies shown are from experimental evaluations of programs operated by governments at a scale of at least 5,000 or more trained. Point estimate of impact of being offered the program and 95 percent confidence intervals shown. Impacts shown are for longest time period post-training for which impact on having paid work and on total labor earnings are available. Percent increase in earnings is relative to control mean. The Colombian and Dominican Republic studies only report results separately by gender, whereas the other studies report results for a pooled sample of men and women.

4.4 percentage points. With the exception of the impact for women of the Colombian *Jóvenes en Acción* program (Attanasio, Kugler, and Meghir 2011), the estimated effect on employment is 2 percentage points or lower in these programs, half that of the meta-analysis impact across all pilot, nongovernment organizations and government programs of 4.0 percentage points reported by Agarwal and Mani (2023). There is more variability in the impact in earnings, but five out of the seven reported estimates are also below the meta-analysis average impact of 8.2 percent.

One plausible reason for the limited effect of these large government programs is that they may not be creating the skills that the labor market is demanding. Some small and informal firms may have no demand for skilled labor at all and lack the physical and managerial capital to benefit from it. For those firms that would like skilled workers, training programs may be slow to update courses and curricula to reflect the changing needs of firms, or perhaps training providers are of poor quality and are incentivized based on number of people trained rather than on employment outcomes. Training may even backfire and cause a reduction in employment if it creates unrealistic expectations among jobseekers, causing them to raise their reservation wages and only search for jobs in the area in which they were trained (Acevedo et al. 2020).

In part due to the evidence from the first waves of rigorous training evaluations, policy efforts have aimed to improve the effectiveness of vocational training programs. The two approaches usually mentioned are to make training more demand-driven and to link payments for the training program more clearly to results. Demand-driven programs aim to have private sector firms and providers, rather than the government, determine what courses are offered and how they are delivered, and to link on-the-job training explicitly to employer demand. An often-mentioned example is the group of *Jóvenes* programs in Latin America, including the Colombian program studied by Attanasio, Kugler, and Meghir (2011), which had the largest impacts among the government programs summarized in Figure 1. Attanasio et al. (2017) link participants in this program to social security records and find, up to a decade later, a lasting effect of 3.8 percentage points on being employed in the formal sector, with trained individuals earning US\$13 more per month in formal earnings. But simply having private sector providers offering the training may be insufficient: Hirshleifer et al. (2016) find privately run courses in Turkey have larger short-term impacts than government-run courses, but that this difference disappears in the three years after training.

Results-based contracting aims to increase the incentives for providers to deliver employment impacts by linking some of the payments to targets, such as the percent of trainees in jobs. World Bank (2020) discusses some of the practical issues and experience with such an approach. It sounds promising in theory, but in practice many governments lack the administrative capacity to measure results and manage such a process. In addition, the share of the total payment linked to performance may in practice end up being relatively small and incentivize only short-term, and not long-term, employment outcomes.

A more optimistic view of the potential impact of job training programs has emerged from impact evaluations of several programs implemented by nongovernment organizations. Particularly influential here has been the work by Alfonsi et al. (2020), who evaluate the impact of vocational training and firm-provided training in the form of apprenticeships in programs operated in Uganda by the nongovernment organization BRAC. The program is much smaller in scale than government programs (697 youth get vocational training and 283 get apprenticeships), relatively intensive (six months duration), with training restricted to a narrow set of sectors

identified as having substantial demand for skilled workers and with a small set of training providers that were selected based on quality. They find the firm apprenticeships have positive short-run impacts that fade out, which they attribute to a lack of skill certification. In contrast, the vocational training has impacts that grow and then stabilize: those assigned to vocational training are 9 percentage points more likely to be employed and earn 25 percent more than the control group averaged over the three years. Shonchoy, Fujii, and Raihan (2018) work with the nongovernment organization Gana Unnayan Kendra in Bangladesh and highlight another way nongovernment organizations may help enhance the effectiveness of training programs: by alleviating other constraints that inhibit youth from using the skills learned. They find that one month of training to work in garment factories has much larger impacts when paired with assistance to migrate to the cities where these jobs are located.

However, while these programs led by nongovernment organizations do offer some lessons for public policy, there are reasons to be cautious in expecting them to be a cost-effective jobs solution at large scale for thousands of jobseekers. First, the impacts of programs tend to fall with scale, which List (2022) dubs the “voltage effect,” in part because of the challenges of ensuring the quality of training is maintained and the topics provided continue to meet the needs of employers at scale. In addition, general equilibrium concerns may arise with scale; that is, training jobseekers en masse in a limited range of skills may result in them all competing with one another for a fixed supply of jobs. Second, the impressive-sounding percent increase in earnings in many of these studies often comes from a relatively small absolute increase in earnings divided by a small base income that disadvantaged individuals would be earning in the absence of training. As a consequence, the gain in income would typically have to last for many more years than studies typically measure in order to pass a cost-benefit test. For example, the 25 percent increase in income in Alfonsi et al. (2020) equates to an extra \$6.10 per month, for a program that costs \$470 per person to provide, and the 16.9 percent increase in income in Crépon and Premand (2021) equates to a \$16.20 per month increase for a program that cost \$2,045 per person. A combined vocational training and life skills program for adolescent girls in Uganda run by the nongovernment organization BRAC had an incredible 308 percent increase in earnings (Bandiera et al. 2020), but this still only equates to an additional \$4.20 per month.

Thus, in many cases the seemingly large percentage earnings gains do not reflect transformational absolute income gains, and will need to persist for five to ten years, or longer, to pass cost-benefit tests. Alternatively, such programs could target poor individuals who alternatively would be receiving even more expensive forms of government support in social assistance programs, which could help them to pass a cost-benefit test.

These modest average impacts of vocational training suggest that a lack of skills is unlikely to be the single binding constraint to finding employment or earning higher wages for the majority of jobseekers taking part in such programs. But modest average impacts may mask large effects for certain subsets of individuals.

An unexplored area for research and policy is testing better ways of targeting the selection of participants into such programs based on those most likely to benefit.

### **Labor Market Intermediation**

Based on the notion that workers may not know *how* or *where* to find available jobs, labor intermediation services seek to equip workers with the tools to improve their job search and to connect workers with jobs. Earlier evidence on traditional labor market intermediation programs, such as government intermediation and placement services, resume and interview preparation, job fairs, and simply sending information about job openings to jobseekers, found that they had only limited and short-term impacts (McKenzie 2017). One possible reason is that these programs may help workers learn to find one job, but given high job turnover, they are little help in finding subsequent work. Similarly, Card, Kluve, and Weber (2018) report job search assistance programs only having average impacts of 1.1 to 2.0 percentage points in their meta-analysis of impacts in high-income countries.

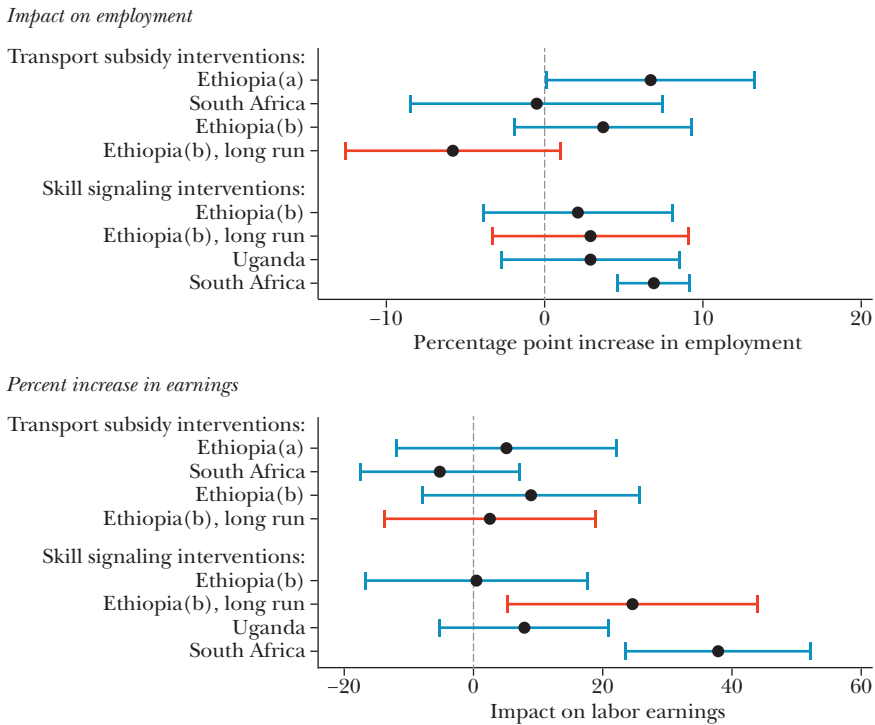
Recent experimental evaluations in developing countries instead show somewhat more promise for interventions that get jobseekers to search in new locations, update biased beliefs, and better signal their skills. Figure 4 summarizes examples of the shorter- and longer-term impacts of two types of these interventions: transport subsidies to help overcome spatial frictions and skill signaling interventions to help reduce information frictions. We discuss these types of interventions in turn, and also mention a relatively new research area—interventions to address behavioral and psychological factors that may be limiting employment.

In developing country labor markets, it is not uncommon to find a surplus of workers relative to available jobs in some locations coexisting with employers in other locations experiencing shortages of similar workers. Even when jobseekers living far from job centers have a job or can easily find one in their local labor markets, the higher quality jobs offering stability, protections, and higher salaries tend to remain out of their reach. For instance, comparing individuals with the same education attainment, Franklin (2018) finds that the share of workers employed informally and in low-skilled occupations increases, and the share employed in high-skilled occupations decreases, with distance to larger cities. Such spatial mismatches are even more striking when considering search across international borders. Jobseekers within an urban area do not seem to search broadly enough in city centers, and jobseekers outside an urban area do not seem to search broadly enough nearby urban areas, despite potentially high returns to searching for jobs over larger distances.

One approach to this issue used by researchers (but not typically by governments at scale) has been to subsidize search across space directly through transportation subsidies. In Ethiopia, Franklin (2018) found youth given these subsidies were more likely to find employment in the city center, and to find jobs of higher quality and permanent jobs, rather than the kinds of casual jobs available in their vicinity. However, as Figure 4 shows, the impacts of such assistance may not last. In Ethiopia,

Figure 4

**Subsidizing Search over Distance Has Positive but Often Temporary Impacts, Whereas Skill Signaling Interventions That Improve Match Quality Can Have More Lasting Impact**



Source: Ethiopia (a) estimates are from Franklin (2018) at four months post-intervention; South Africa estimates from Banerjee and Sequeira (2020) at one year post-intervention; Ethiopia (b) estimates from Abebe et al. (2021) at one year and four years post-intervention; Uganda estimates from Bassi and Nansamba (2022) at one year post-intervention; South Africa estimates from Carranza et al. (2022) at four months post-intervention.

Notes: Point estimate of experimental impacts and 95 percent confidence intervals shown of being offered the program. Impacts shown are for longest time period post-training for which impact on having paid work and on total labor earnings are available. Percent increase in earnings is relative to control mean.

Abebe et al. (2021) find a modest impact of transport subsidies on permanent and formal employment, with little to no improvement in jobseekers' probability of having a job four years after this support is withdrawn.

Distance to jobs and high commuting costs could in part explain the declining impacts over time. Job quits may occur if jobseekers initially underestimate the disutility of commuting long distances and this offsets the wage premium paid in larger cities and urban labor markets (Banerjee and Sequeira 2020). When poor

matches between workers and firms result in high rates of job turnover, jobseekers recurrently have to search for jobs and may again find difficult to access better opportunities in distant labor markets without a repeated subsidy.

A one-time transport subsidy can lead to longer-run effects on employment if searching more broadly allows jobseekers to learn about the spatial distribution of wages, and build job connections in a wider labor market. Strong repeated search is similarly possible if the spatial wage premium is large, and jobseekers can build assets over time as a result of obtaining employment in the short-term. An example comes from the work of Bryan, Chowdhury, and Mobarak (2014), who provided financial assistance to help subsistence rural households in Bangladesh migrate for work in nearby urban areas during the agricultural lean season. Not only did this result in better jobs in the short-run, but once workers learn about and experience the benefits of jobs in urban areas, they return for work in following years without further incentive.

Subsidies to address spatial frictions are likely to work best for that subset of jobseekers for whom cash constraints, lack of experience, and lack of networks is the most important and binding constraint to participating in jobs in this new location. Mitchell et al. (2023) report that efforts to scale the Bangladeshi program resulted in much lower impacts, because financial assistance ended up largely going to people who were inclined to migrate anyway. But the very poor and disadvantaged may find it harder to save and afford the costs of repeat travel, so that there will also be no long-term impacts for this group. Hence, while spatial frictions are important, transport subsidies to overcome these frictions need careful targeting.

An alternative set of interventions to improve labor market intermediation seeks to address information frictions, overcome biased beliefs, and signal skills. Remember, most jobs in developing countries are found through social networks and direct contact with employers. Thus, many workers may have inaccurate beliefs about the full range of job opportunities and wages available. This concern may be especially salient for young, racial minority, and poorly-educated workers, who may be segregated from networks that provide information and contacts on many better jobs. The result of these biased beliefs can be that jobseekers may not search for jobs that could be a good match for them. Such workers may also have too high a reservation wage (Alfonsi, Namubiru, and Spaziani 2022), or reservation job prestige (Groh et al. 2015), causing them to turn down jobs they could get, choose poorly matched jobs, and quit soon after starting.

Can new, valuable, credible information cause an updating of beliefs, and in this way result in employment and earnings gains? In Uganda, Alfonsi, Namubiru, and Spaziani (2022) find trainees in a vocational education program overestimate how much they will earn in their first job, resulting in high reservation wages, but also underestimate the returns to experience and salary growth potential possible after starting work. Mentors who had been through the same program several years earlier were able to credibly help jobseekers form more realistic expectations, causing them to revise reservation wages downwards, turn down fewer jobs, and earn 18 percent more a year later.

In developing countries, information about the skills of workers can pose frictions to hiring. Potential employers find it difficult to assess the ability of workers, especially those with low levels of education. Jobseekers may have limited ability to know their own job skills and how they compare to other candidates, which can affect their job search behavior. To the extent that employers are more uncertain or underestimate the ability of lower-educated jobseekers and jobseekers who have weaker labor market links—racial minority, younger, and female workers—information frictions can also exacerbate labor market inequality.

Enabling jobseekers to communicate their skills in a credible manner can help overcome this information friction, in this way increasing the likelihood of workers getting hired and helping firms to select better candidates. Figure 4 shows that these types of interventions can generate more lasting impacts. In Ethiopia, Abebe et al. (2021) implement job application workshops that test and certify jobseekers on their skills and teach them how to signal these skills to employers. This intervention sped up entry into formal jobs, and treated youth have 25 percent higher earnings four years later, in contrast to the only temporary earnings impact of transport subsidies on the same population. In South Africa, helping jobseekers signal cognitive and noncognitive skills assessment results to firms increased their employment rate and earnings (Carranza et al. 2022). In Uganda, providing certificates of soft skills led workers who found employment to earn more (Bassi and Nansamba 2022). Helping jobseekers obtain and understand the benefits of references from previous employers is another way of helping signal skills. Abel, Burger, and Piraino (2020) found that reference letters lead firms to select candidates of higher ability (as captured by test scores not shared with firms), suggesting that assessment of workers' past performance conveys additional information to employers, improving their capacity to screen better applicants.

Of course, such skill certification will only work if these information search frictions are a main reason why workers are not receiving job offers they would like. In Jordan, Groh et al. (2015) implemented testing and skill certification with no impact on employment—which they attribute to much of the unemployment being driven by strong preferences over nonwage job attributes.

Finally, even if jobseekers know where and how to look for jobs, and can credibly signal their skills, behavioral and psychological factors may still limit job search intensity and success rates. Certain low-cost tweaks can take these factors into account when designing other interventions. This area is a relatively new, but we provide three examples that indicate the types of interventions that show promise. First, as a result of the so-called “intention-behavior gap,” jobseekers may fail to follow through in submitting as many job applications as they would ideally prefer. An intervention in South African Labor Centers that prompted jobseekers to set plans and goals for job search increased job search intensity, resulting more job offers and higher employment over the next three months (Abel et al. 2019). Second, workers who find a job may not always show up on time or comply with workplace rules, leading to high turnover. A soft-skills training program focused on activating conscientiousness reduced job turnover among construction workers in Senegal

(Allemand et al. 2023). Finally, high discount rates and impatience may mean that even if workers recognize the returns to experience, they may be unwilling to accept jobs with relatively low starting wages and high wage growth trajectories. In a study in Mexico, Abel et al. (2022) find that a temporary wage subsidy can help overcome this behavioral bias and increase formal employment rates as a result.

### **Online Job Platforms**

The number of workers and firms using online jobs platforms has undoubtedly been rising as internet penetration has increased in developing countries and new platforms have launched. For example, the Nigerian platform Jobberman claims 2.6 million jobseekers and 75,000 employers on its platform (as reported in Ladipo 2022), while the Indian portal Naukri has an estimated database of 82 million job seekers and 5 million recruiters, with an estimate of over 7 million searches by recruiters conducted daily.<sup>7</sup> These online portals can lower the costs of search, enable search across space, and help alleviate information frictions, making it easier for job matches to occur. Apparently, many users believe these platforms are at least somewhat beneficial.

The rise of online job portals raises two questions for public policy. First, should governments be trying to set up additional job portals beyond what the private sector is already doing? It is unclear that governments have a comparative advantage in doing this for most types of jobs. Indeed, governments may be worse at performing ongoing platform maintenance. Governments may also face policy restrictions on the types of jobs they can post; for example, only posting formal jobs that comply with safety and nondiscrimination rules, in labor markets with a high share of informal jobs. Some attempted government efforts have not performed well. For example, Ghana's Youth Employment Agency set up an online jobs portal (<https://yeajobcentre.gov.gh/>) in November 2019. In March 2023, the entire portal advertised only 60 jobs, in a country of over 32 million people. Our view is that the most compelling case for governments to set up portals are for specialized sectors in which the government plays a strong role in hiring—say, for public employment in health and education—or potentially when setting up new bilateral international migration agreements. But even in these cases, the government should probably look for private sector firms with a viable business model for operating these portals.

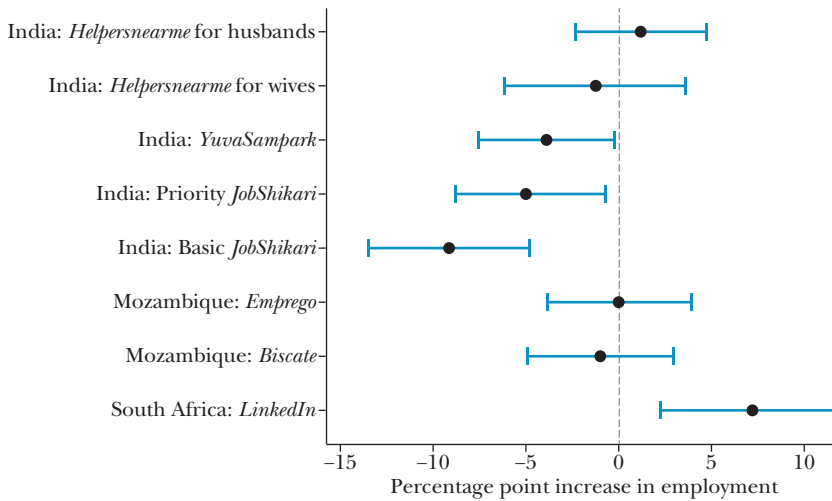
Second, even if the government is not setting up its own portals, should it encourage jobseekers to use such portals? Figure 5 summarizes the results of five recent randomized experiments that provided nudges or assistance to encourage a treatment group to start using an existing platform. In four of these cases, the interventions had minimal or even negative impacts on the likelihood of jobseekers finding employment. One possible reason is that many of those encouraged to use the services do not take them up. Alternatively, Kelley, Ksoll, and Magruder (2022) suggest that jobseekers may be overoptimistic in their expectations of the jobs

<sup>7</sup>Estimates here are from the business management software company Freshworks at <https://www.freshworks.com/hrms/indeed-vs-naukri-choose-the-best-prescreening-tool/> (accessed March 12, 2023).



Figure 5

### Most Efforts Encouraging Jobseekers to Use Online Job Portals Have Not Significantly Boosted Employment



Source: India *Helpersnearme* results are from Afridi et al. (2022), who encouraged husbands and wives to register on a local blue collar job platform; India *YuvaSampark* results are from Chakravorty et al. (2023), who encouraged vocational training graduates to register for a government jobs app; India *JobShikari* results are from Kelley, Ksoll, and Magruder (2022), who enrolled vocational training graduates in a platform that sends SMS messages about relevant openings, with a priority group getting a higher rank and more messages; Mozambique results are from Jones and Sen (2022), who encouraged graduates of technical and vocational colleges to sign up for two local jobs platforms: *Biscate* (informal manual jobs) and *Emprego* (formal jobs); South Africa results are from Wheeler et al. (2022), who trained young, disadvantaged youth who had participated in a job readiness program how to use *LinkedIn*.

Note: Point estimate and 95 percent confidence interval shown of experimental impacts of offering use of online job portals.

they can find on such a platform, and take time to adjust their reservation wages downwards. Selection issues may arise here, as well: perhaps the compliers in these experiments—that is, individuals who only use the platform if nudged—are a separate group from those who benefit, using such platforms without any policy assistance and thus are not included in these studies. An exception to these disappointing impacts is Wheeler et al. (2022), who find that training disadvantaged South African youth who had been through a job readiness program to use LinkedIn increases their employment by 6.9 percentage points over the next year. This finding may reflect several factors: perhaps the training provided more information to jobseekers than just encouragement to use the platform; perhaps this program was better targeted at marginal, disadvantaged jobseekers; or perhaps this segment of the South African labor market had more information frictions than some other labor markets.

Government policy might also seek to improve the functioning of these job portals. One problem is fragmentation; that is, a proliferation of job portals can make it more difficult for jobseekers and employers to find one another. A potential solution is for government employment agencies to work as aggregators of vacancy information from different platforms, as is done in Colombia. Another problem is trust issues; after all, firms have often relied on personal connections and networks for hiring in part to overcome trust issues. In a study of small firms in India, Fernando, Singh, and Tourek (forthcoming) find that offering verification of skills along with an expanded pool of candidates makes these firms more likely to hire workers on an online platform. While platforms themselves can provide some verification services, government education and training programs can also do this, and government can also play a role through criminal background checks, credit records, and other reputation mechanisms.

## **Conclusion**

Employment in developing countries is often dominated by small and informal firms, and thus faces a shortage of formal-sector jobs with good wages. Job training and job search policies by themselves are unlikely to generate a lot of new employment, and there is a need for complementary policies that aim to spur the demand for labor. Nevertheless, there is still a role for well-designed policies to help speed up the process of structural transformation in labor markets in ways that will improve employer-employee matches and thus increase the productivity and wages of available jobs, while also improving the employment prospects of disadvantaged groups.

However, governments often struggle to implement these job training and job search policies successfully at scale. How might these policies be implemented more effectively? We have discussed some general principles, but much depends on tailoring solutions to meet the needs of specific localities, sectors, and types of jobseekers—which is another reason that centralized national programs often struggle. For job training, programs need to be closely tied to market demand, so that employers want to hire workers who are trained. The returns to training to the average jobseeker in a large-scale government program are typically very modest, and much more work is needed to determine how best to target training programs and how to expand successful pilot programs to larger scale. Job search assistance seems to work best when it helps jobseekers learn not just about a particular job, but rather learn something more fundamental that causes them to update their beliefs about the types of jobs they should be considering (including which sectors and locations to search in), and the wage levels and trajectories in those jobs. Efforts to certify the existing skills of workers in a credible manner can help reduce information frictions when education systems do not provide good signals. Performing these tasks well will require investing in good data systems—otherwise, labor markets in which it is difficult for workers, and firms failing to create lasting and well-paid jobs, will also be challenging for governments to understand.

■ *We are grateful to the editors, Heidi Williams, Erik Hurst, Nina Pavcnik, and Timothy Taylor, for their useful comments and their help in improving this piece. We thank Yves Perardel for providing us with tabulations from the Albania and Romania labor force surveys.*

## References

- Abebe, Girum, A. Stefano Caria, Marcel Fafchamps, Paolo Falco, Simon Franklin, and Simon Quinn. 2021. "Anonymity or Distance? Job Search and Labour Market Exclusion in a Growing African City." *Review of Economic Studies* 88 (3): 1279–1310.
- Abebe, Girum, Stefano Caria, Marcel Fafchamps, Paolo Falco, Simon Franklin, Simon Quinn, and Forhad Shilpi. 2020. "Matching Frictions and Distorted Beliefs: Evidence from a Job Fairs Experiment." Unpublished.
- Abel, Martin, Rulof Burger, Eliana Carranza, and Patrizio Piraino. 2019. "Bridging the Intention-Behavior Gap? The Effect of Plan-Making Prompts on Job Search and Employment." *American Economic Journal: Applied Economics* 11 (2): 284–301.
- Abel, Martin, Rulof Burger, and Patrizio Piraino. 2020. "The Value of Reference Letters: Experimental Evidence from South Africa." *American Economic Journal: Applied Economics* 12 (3): 40–71.
- Abel, Martin, Eliana Carranza, Kimberly Geronimo, and Maria Elena Ortega. 2022. "Can Temporary Wage Incentives Increase Formal Employment? Experimental Evidence from Mexico." IZA Discussion Paper 15740.
- Acevedo, Paloma, Guillermo Cruces, Paul Gertler, and Sebastian Martinez. 2020. "How Vocational Education Made Women Better Off but Left Men Behind." *Labour Economics* 65: 101824.
- Afridi, Farzana, Amrita Dhillon, Sanchari Roy, and Nikita Sangwan. 2022. "Social Networks, Gender Norms and Women's Labor Supply: Experimental Evidence Using a Job Search Platform." IZA Discussion Paper 15767.
- Agarwal, Neha, and Subha Mani. Forthcoming. "New Evidence on Vocational and Apprenticeship Training Programs in Developing Countries." In *The Handbook of Experimental Development Economics*, edited by Utteeyo Dasgupta and Pushkar Maitra. Cheltenham: Edward Elgar Publishing Ltd.
- Alfonsi, Livia, Oriana Bandiera, Vittorio Bassi, Robin Burgess, Imran Rasul, Munshi Sulaiman, and Anna Vitali. 2020. "Tackling Youth Unemployment: Evidence from a Labor Market Experiment in Uganda." *Econometrica* 88 (6): 2369–2414.
- Alfonsi, Livia, Mary Namubiru, and Sara Spaziani. 2022. "Meet Your Future: Experimental Evidence on the Labor Market Effects of Mentors." Unpublished.
- Allemand, Mathias, Martina Kirchberger, Sveta Milusheva, Carol Newman, Brent Roberts, and Vincent Thorne. 2023. "Conscientiousness and Labor Market Returns: Evidence from a Field Experiment in West Africa." Trinity Economics Working Paper 0123.
- Assaad, Ragui. 2014. "Making Sense of Arab Labor Markets: The Enduring Legacy of Dualism." *IZA Journal of Labor and Development* 3: 6.
- Attanasio, Orazio, Arlen Guarín, Carlos Medina, and Costas Meghir. 2017. "Vocational Training for Disadvantaged Youth in Colombia: A Long-Term Follow-Up." *American Economic Journal: Applied Economics* 9 (2): 131–43.
- Attanasio, Orazio, Adriana Kugler, and Costas Meghir. 2011. "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics* 3 (3): 188–220.
- Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2020. "Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa." *American Economic Journal: Applied Economics* 12 (1): 210–59.
- Banerjee, Abhijit, and Sandra Sequeira. 2020. "Learning by Searching: Spatial Mismatches and Imperfect

- Information in Southern Labor Markets.” *Journal of Development Economics* 164: 103111.
- Bassi, Vittorio, and Aisha Nansamba.** 2022. “Screening and Signaling Non-cognitive Skills: Experimental Evidence from Uganda.” *Economic Journal* 132 (642): 471–511.
- Bertrand, Marianne, and Bruno Crépon.** 2021. “Teaching Labor Laws: Evidence from a Randomized Control Trial in South Africa.” *American Economic Journal: Applied Economics* 13 (4): 125–49.
- Bertand, Marianne, Chang-Tai Hsieh, and Nick Tsivanidis.** 2022. “Contract Labor and Establishment Growth in India.” Unpublished.
- Bhorat, Haroon, Adaiah Lilenstein, and Ben Stanwix.** 2021. “The Impact of the National Minimum Wage in South Africa: Early Quantitative Evidence.” University of Cape Town DPRU Working Paper 202104.
- Blattman, Christopher, and Stefan Dercon.** 2018. “The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia.” *American Economic Journal: Applied Economics* 10 (3): 1–38.
- Blattman, Christopher, and Laura Ralston.** 2015. “Generating Employment in Poor and Fragile States: Evidence from Labor Market and Entrepreneurship Programs.” Unpublished.
- Brown, Gabriel, Morgan Hardy, Isaac Mbiti, Jamie McCasland, and Isabelle Salcher.** Forthcoming. “Can Financial Incentives to Firms Improve Apprentice Training? Experimental Evidence from Ghana.” *American Economic Review: Insights*.
- 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.
- Caicedo, Santiago, Miguel Espinosa, and Arthur Seibold.** 2022. “Unwilling to Train? Firm Responses to the Colombian Apprenticeship Regulation.” *Econometrica* 90 (2): 507–50.
- Card, David, Jochen Kluge, and Andrea Weber.** 2018. “What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations.” *Journal of the European Economic Association* 16(3): 894–931.
- Carranza, Eliana, Robert Garlick, Kate Orkin, and Neil Rankin.** 2022. “Job Search and Hiring with Limited Information about Workseekers’ Skills.” *American Economic Review* 112 (11): 3547–83.
- Carranza, Eliana, and David McKenzie.** 2023. “Reproducibility Package for Job Training and Job Search Assistance Policies in Developing Countries.” World Bank Reproducible Research Repository. <https://doi.org/10.60572/HM2R-NJ04>.
- Chakravorty, Bhaskar, Apurav Yash Bhatiya, Clément Imbert, Maximilian Lohnert, Poonam Panda, and Roland Rathelot.** 2023. “Impact of the COVID-19 Crisis on India’s Rural Youth: Evidence from a Panel Survey and an Experiment.” *World Development* 168: 106242.
- Chen, Yawen.** 2022. “This Chinese Jobs Crisis Could Be Its Worst.” *Reuters*, June 20. <https://www.reuters.com/breakingviews/this-chinese-jobs-crisis-could-be-its-worst-2022-06-20/>.
- Crépon, Bruno, and Patrick Premand.** 2019. “Direct and Indirect Effects of Subsidized Dual Apprenticeships.” IZA Discussion Paper 12793.
- de Mel, Suresh, David McKenzie, and Christopher Woodruff.** 2019. “Labor Drops: Experimental Evidence on the Return to Additional Labor in Microenterprises.” *American Economic Journal: Applied Economics* 11 (1): 202–35.
- Donkor, Audrey.** 2021. “Africa’s Youth Unemployment Crisis Is a Global Problem.” *Foreign Policy*, October 19. <https://foreignpolicy.com/2021/10/19/africa-youth-unemployment-crisis-global-problem/>.
- Fernando, A. Nilesh, Niharika Singh, and Gabriel Tourek.** Forthcoming. “Hiring Frictions and the Promise of Online Job Portals: Evidence from India.” *American Economic Review: Insights*.
- Franklin, Simon.** 2018. “Location, Search Costs and Youth Unemployment: Experimental Evidence from Transport Subsidies.” *Economic Journal* 128 (614): 2353–79.
- Gehrke, Esther, and Renate Hartwig.** 2018. “Productive Effects of Public Works Programs: What Do We Know? What Should We Know?” *World Development* 107: 111–24.
- Groh, Matthew, Nandini Krishnan, David McKenzie, and Tara Vishwanath.** 2013. “Jordan New Work Opportunities for Women Pilot Impact Evaluation (NOW) 2010-2013, Ref. JOR\_2010\_JNOW\_v01\_M\_v01\_A\_PUF.” Dataset downloaded from <https://microdata.worldbank.org/index.php/catalog/1549> on August 31, 2023.
- Groh, Matthew, David McKenzie, Nour Shammout, and Tara Vishwanath.** 2015. “Testing the Importance of Search Frictions and Matching through a Randomized Experiment in Jordan.” *IZA Journal of Labor Economics* 4: 7.
- Hardy, Morgan, Isaac Mbiti, Jamie McCasland, and Isabelle Salcher.** 2019. “The Apprenticeship-to-Work Transition: Experimental Evidence from Ghana.” World Bank Policy Research Working Paper 8851.
- Hirshleifer, Sarojini, David McKenzie, Rita Almeida, and Cristobal Ridao-Cano.** 2016. “The Impact of

- Vocational Training for the Unemployed: Experimental Evidence from Turkey.” *Economic Journal* 126 (597): 2115–46.
- ILOSTAT**. 2019. “Albania Labour Force Survey 2019.” <https://www.ilo.org/surveyLib/index.php/catalog/7902/study-description> (accessed on August 31, 2023).
- ILOSTAT**. 2021. “Romania Household Labor Force Survey 2021.” <https://www.ilo.org/surveyLib/index.php/catalog/8046/related-materials> (accessed August 31, 2023).
- INEGI**. 2023. “Encuesta Nacional de Ocupación y Empleo (ENOE) (National Survey of Occupation and Employment).” <https://www.inegi.org.mx/programas/enoe/14ymas/#Microdatos> (accessed on August 31, 2023).
- Jones, Sam, and Kunal Sen**. 2022. “Labour Market Effects of Digital Matching Platforms: Experimental Evidence from Sub-Saharan Africa.” WIDER Working Paper 2022/69.
- Kelley, Erin M., Christopher Ksoll, and Jeremy Magruder**. 2022. “How Do Online Job Portals Affect Employment and Job Search? Evidence from India.” Unpublished.
- Ladipo, Ebuloluwa**. 2022. “Jobberman places 130,000 Nigerians in Paid Employment in 2 Years.” *Business Day* (Nigeria), January 25. <https://businessday.ng/news/article/jobberman-places-130000-nigerians-in-paid-employment-in-2-years/> (accessed March 12, 2023).
- List, John A.** 2022. *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale*. New York: Currency.
- Maitra, Srabani, Saikat Maitra, and Manish Thakur**. 2022. “Uncertain Itineraries: Dual System of Training and Contemporary TVET Reforms in India.” *Journal of Vocational Education and Training*. <https://doi.org/10.1080/13636820.2022.2042724>.
- Manpower Group**. 2023. “Manpower Group Employment Outlook Survey Q2 2023.” [https://go.manpowergroup.com/hubfs/GLOBAL\\_EN\\_Report\\_2Q23.pdf](https://go.manpowergroup.com/hubfs/GLOBAL_EN_Report_2Q23.pdf).
- McKenzie, David**. 2017. “How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence.” *World Bank Research Observer* 32 (2): 127–54.
- McKenzie, David, Rita Almeida, Cristobal Ridao-Cano, Sarojini Hirschleifer**. 2014. “Vocational Training Program for the Unemployed Impact Evaluation 2010-2012. Ref. TUR\_2010\_VTPUIE\_v01\_M.” Dataset downloaded from <https://microdata.worldbank.org/index.php/catalog/1973> on August 31, 2023.
- McKenzie, David, Christopher Woodruff, Kjetil Bjorvatn, Miriam Bruhn, Jing Cai, Juanita Gonzalez-Uribe, Simon Quinn, Tetsushi Sonobe, and Martin Valdivia**. 2021. “Training Entrepreneurs.” *VoxDevLit* 1 (2): 1–27.
- Middle East Monitor**. 2023. “Jordan Unemployment Rate Hits 23.1%.” January 6. <https://www.middleeastmonitor.com/20230106-jordan-unemployment-rate-hits-23-1/>.
- Mitchell, Harrison, A. Mushfiq Mobarak, Karim Naguib, Maira Emy Reimão, and Ashish Shenoy**. 2023. “Delegation Risk and Implementation at Scale: Evidence from a Migration Loan Program in Bangladesh.” Unpublished.
- NewsIn.Asia**. 2023. “‘India Is Sitting on a Time Bomb’: Jobs Crisis Looms as Population Soars.” January 18. <https://newsin.asia/india-is-sitting-on-a-time-bomb-jobs-crisis-looms-as-population-soars/>.
- Ridgwell, Henry**. 2018. “Robots Pose Big Threat to Jobs in Africa, Researchers Warn.” *Voice of America*, March 28. <https://www.voanews.com/a/robots-in-africa/4320405.html>.
- Shonchoy, Abu, Tomoki Fujii, and Selim Raihan**. 2018. “Barriers to Labor Migration for the Rural Poor: Experimental Evidence from a Vocational Training Program in Bangladesh.” Unpublished.
- Statistics Sierra Leone**. 2014. “Sierra Leone Labor Force Survey (LFS) 2014. Ref. SLE\_2014\_LFS\_v01\_M.” Dataset downloaded from <https://microdata.worldbank.org/index.php/catalog/2687> on August 31, 2023.
- Taylor, Lin**. 2019. “Rising Heat Stress Could Cost 80 Million Jobs by 2030, UN Says.” Thomas Reuters Foundation, July 2. <https://www.globalcitizen.org/en/content/rising-heat-stress-cost-jobs/>.
- Wheeler, Laurel, Robert Garlick, Eric Johnson, Patrick Shaw, and Marissa Gargano**. 2022. “LinkedIn(to) Job Opportunities: Experimental Evidence from Job Readiness Training.” *American Economic Journal: Applied Economics* 14 (2): 101–25.
- World Bank**. 2013. “Morocco Household and Youth Survey (MHYS) 2009-2010, Ref. MAR\_2009\_MHYS\_v01\_M\_PUF.” Dataset downloaded from <https://microdata.worldbank.org/index.php/catalog/1546> on August 31, 2023.
- World Bank**. 2020. *Demand-Driven Skills Training and Results-Based Contracting: Lessons for Youth Employment Programs*. World Bank: Washington, DC.
- World Bank**. 2023. “World Development Indicators Databank.” <https://data.worldbank.org/indicator/> (accessed on August 31, 2023).



## 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., Saint Paul, MN 55105.*

### Smorgasbord

Claudia Goldin has been awarded the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2023, “for having advanced our understanding of women’s labour market outcomes.” The Nobel committee publishes useful background essays and videos at <https://www.nobelprize.org/prizes/economic-sciences/2023/summary/>. From the “Popular Science Background”: “Goldin began by compiling statistics from a range of sources, producing the first long series on the pay gap between men and women. Using materials that covered two hundred years, she was able to demonstrate that many historically important structural changes in the labour market actually benefitted women, long before the issue of equality

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.38.1.245>.

was a priority. The gender earnings gap lessened significantly during the industrial revolution (1820–1850) and when demand for administrative and clerical services increased (1890–1930). However, despite economic growth, increasing education levels among women and a doubling in the proportion of women working for pay, the earnings gap essentially stayed the same between 1930 and 1980. . . . [T]he earnings gap between women and men in high-income countries is somewhere between ten and twenty per cent, even though many of these countries have equal pay legislation and women are often more educated than men. Why is this? Goldin attempts to answer precisely this question and, among other things, succeeds in identifying one key explanation: parenthood. . . . Goldin showed that this motherhood effect can partly be explained by the nature of contemporary labour markets, where many sectors expect employees to be constantly available and flexible in the face of the employer’s demands. Because women often take greater responsibility than men for childcare, for example, this makes career progression and earnings increases more difficult.”

The *Annual Review of Financial Economics* has published a four-paper symposium, with an introduction, on the subject of choosing an appropriate social discount rate (2023, <https://www.annualreviews.org/toc/financial/15/1>). From the abstract of the paper by W. Kip Viscusi, “The Social Discount Rate: Legal and Philosophical Underpinnings”: “Discounting of deferred impacts of government policies is a long-established practice that has been the target of substantial litigation and continued philosophical debate. Legal challenges to the social rate of discount have resulted in general acceptance of the principle of discounting at a nonzero rate for both monetary and nonmonetary impacts. Courts have displayed a general familiarity with discounting and often require transparent justification for the selection of the discount rate based on established scientific principles. The philosophical issues are more wide-ranging and include whether nonmonetary impacts should be discounted, the use of the opportunity cost of capital or the social rate of time preference, the time frame that is pertinent for setting the discount rate, and determination of whose preferences should have standing. Intergenerational issues are particularly challenging, raising questions regarding which generation’s preferences should be recognized, the potential for dynamic inconsistencies arising from preferential discount rates, and intertemporal inequities.”

Lawrence H. Summers delivered the first annual lecture honoring Richard N. Cooper, at the Peterson Institute for International Economics, with the title “What should the 2023 Washington Consensus be?” (September 18, 2023, video and transcript available, <https://www.piie.com/events/what-should-2023-washington-consensus-be>). Summers organizes his discussion around “six major misconceptions.” Here are two of them: “First, it is supposed that the idea of economic policy is to maximize the creation of jobs rather than to maximize the availability of goods at low cost to consumers and firms. Both the officials responsible for competition policy and those responsible for international trade have explicitly rejected economic efficiency as a central guide for economic policy. This, I would suggest, is a costly and consequential error. . . . Third, the world has fared



very well. Relative to the time when I was chief economist of the World Bank at the beginning of the 1990s, child mortality rates are less than half of what they were then. Literacy rates are more than twice what they were then. Poverty rates, terms of extreme poverty are less than 40% of what they were then. And in some ways most fundamental and important, this month, we celebrate the 78th anniversary of a situation where there has been no direct war between major powers. You cannot find a period of 78 years since Christ was born when that was the case. So, the idea that we've been doing it all wrong is, I would suggest, a substantial misconception."

Mario Draghi delivered the 15th Annual Martin Feldstein Lecture on "The Next Flight of the Bumblebee: The Path to Common Fiscal Policy in the Eurozone" (*NBER Reporter*, October 2023, <https://www.nber.org/reporter/2023number3/next-flight-bumblebee-path-common-fiscal-policy-eurozone>). "Whichever route we take [in the European Union], we cannot stand still or—like a bicycle—we will fall over. The strategies that had ensured our prosperity and security in the past—reliance on the USA for security, on China for exports, and on Russia for energy—are insufficient, uncertain, or unacceptable. The challenges of climate change and migration only add to the sense of urgency to enhance Europe's capacity to act. We will not be able to build that capacity without reviewing Europe's fiscal framework, and I have tried to outline the directions this change might take. But ultimately the war in Ukraine has redefined our Union more profoundly—not only in its membership, and not only in its shared goals, but also in the awareness it has created that our future is entirely in our hands, and in our unity."

There are two broad models for how to address the chronic homeless, which go under the headings of "treatment first" and "housing first." Joseph R. Downes makes the case for the second in "Housing First: A Review of the Evidence" (*Evidence Matters: US Department of Housing and Urban Development*, Spring/Summer 2023, pp. 11–19, <https://www.huduser.gov/portal/periodicals/em/spring-summer-23/highlight2.html>). "Four major RCTs [randomized control trials] have been performed to compare the effectiveness of Housing First programs with treatment first programs. Three of these RCTs were conducted in the United States, and the other was conducted in Canada. In a review of these RCTs, Tsai notes that two RCTs conclusively found that Housing First led to quicker exits from homelessness and greater housing stability than did TAU [treatment as usual]. In the Canadian trial, an RCT in five of Canada's largest cities known as At Home/Chez Soi, analysis revealed that, in findings similar to those of the American RCTs, 'Housing First participants spent 73% of their time in stable housing compared with 32% of those who received treatment as usual.' Baxter et al. also performed a systematic literature review and metanalysis of these four RCTs, finding that Housing First resulted in significant improvements in housing stability. This study also found that no clear differences existed between Housing First and TAU for mental health, quality of life, and substance use outcomes . . ."

The World Trade Organization has published its *World Trade Report 2023* on the theme "Reglobalization for a Secure, Inclusive, and Sustainable Future" (<https://>

[www.wto.org/english/res\\_e/publications\\_e/wtr23\\_e.htm](http://www.wto.org/english/res_e/publications_e/wtr23_e.htm)). From the Executive Summary: “Recent crises, such as the COVID-19 pandemic and the war in Ukraine, have fed into perceptions that globalization exposes economies to excessive risks. Consequently, a trade-sceptic narrative has gained traction, suggesting that international trade is an obstacle to building a more secure, inclusive, and sustainable world. Viewing interdependence as a vice rather than a virtue, policymakers are now placing greater emphasis on economic independence. . . . The primary conclusion of the Report is that international trade, anchored in a strengthened multilateral trading system, plays an indispensable role in creating a more secure, inclusive, and sustainable world. Building upon these findings, the Report makes the case that a better alternative to fragmentation is ‘re-globalization’—understood as extending trade integration to more people, economies and issues.”

## Country Narratives

A research group at Harvard’s Growth Lab has published “Growth Through Inclusion in South Africa” (November 15, 2023, <https://growthlab.hks.harvard.edu/publications/growth-through-inclusion-south-africa>). The authors are Ricardo Hausmann, Tim O’Brien, Andrés Fortunato, Alexia Lochmann, Kishan Shah, Lucila Venturi, Sheyla Enciso-Valdivia (LSE), Ekaterina Vashkinskaya (LSE), Ketan Ahuja, Bailey Klinger, Federico Sturzenegger, and Marcelo Tokman. “It is unfortunately clear that South Africa’s trajectory is not one of growth or inclusion, but rather stagnation and exclusion. South Africa’s economy is stagnating and, in fact, losing capabilities, export diversity, and competitiveness. While the racial composition of wealth at the top has changed, wealth concentration in South Africa has not and remains very high. Moreover, the broader structures of the economy have not allowed for the inclusion of the labor and talents of South Africans—black, white, and otherwise. There appear to be major spatial impediments to labor market inclusion in cities and large spatial patterns of exclusion in former homelands. As the performance of network industries and public capabilities have deteriorated and growth has slowed, exclusion has only worsened. Empowerment of a few has de facto come at the expense of the many.”

Johan Norberg discussed “The Mirage of Swedish Socialism: The Economic History of a Welfare State” (Fraser Institute, August 2023, <https://www.fraserinstitute.org/studies/mirage-of-swedish-socialism-the-economic-history-of-a-welfare-state>). “Sweden has a tradition of sticking to the path it has chosen and ignoring problems until they become too big to deny and everybody changes their minds at the same time. Then Swedes move fast in the opposite direction. Far from following the famed ‘middle way,’ Sweden has often been a country of extremes. It liberalized the economy more than other countries did in the mid-1800s, socialized more than others in the mid-1900s, and then reversed course and liberalized again faster than others in the late 20th century.”

*Ending Stagnation: A New Economic Strategy for Britain* has been published by the Resolution Foundation and the Centre for Economic Policy Research (December 2023, <https://economy2030.resolutionfoundation.org/wp-content/uploads/2023/12/Ending-stagnation-final-report.pdf>). “We [that is, the United Kingdom] were catching up with more-productive countries like France, Germany and the US during the 1990s and early 2000s. But that came to an end in the mid-2000s and our relative performance has been declining ever since, reflecting a productivity slowdown far surpassing those seen in similar economies. Labour productivity grew by just 0.4 per cent a year in the UK in the 12 years following the financial crisis, half the rate of the 25 richest OECD countries (0.9 per cent). The UK’s productivity gap with France, Germany and the US has doubled since 2008 to 18 per cent . . . Weak productivity growth has fed directly into flatlining wages and sluggish income growth: real wages grew by an average of 33 percent a decade from 1970 to 2007, but this fell to below zero in the 2010s . . . We might like to think of ourselves as a country on a par with the likes of France and Germany, but we need to recognise that, except for those at the top, this is simply no longer true when it comes to living standards.”

## Interviews with Economists

Mitchell List and Kurt Schuler present “An Interview with James D. Gwartney on His Life and Work in Economics” (Johns Hopkins Institute for Applied Economics, Global Health and the Study of Business Enterprise, SAE #238, August 2023, <https://www.dropbox.com/scl/fi/0emygx0129345gnbw837w/Working-Paper-238.pdf>). Gwartney talks about attending a one-room schoolhouse in Kansas, as did Vernon Smith. He also discusses the philosophy behind his popular introductory textbook: “I wanted to integrate public choice analysis into a principles of economics text. Around 45 percent of [US] GDP was allocated through the political process in 2020 and through most of the history of the book it’s been in the 30 to 35 percent range. . . . The political process is merely an alternative form of making decisions. We need to know something about how that process works as well as how markets work. This is the contribution of our text. Merely stating, ‘Here’s what the benevolent, omnipotent dictator (an expression my friend Randy Holcombe likes to use when talking about the political process) would do’ is not very useful. Political decisionmakers may not be very benevolent, but even if they are benevolent, they’re not going to be omniscient, therefore there’s no reason to expect that they’re going to come up with ideal solutions. Even today, much of economics reflects this misleading view. Our book integrating public choice was really an attack on the idea that government is a corrective device that’s lying around so that if something goes wrong, we’ll just call on the corrective device and fix it. That seemed a very naive view of what the role of government in the economy should be. I believe this integration of public choice accounts for the staying power of our text.”

Tim Besley serves as interlocutor in “Biodiversity: A Conversation with Sir Partha Dasgupta” (*Annual Review of Economics*, 2023, pp. 755–773, transcript and video available, <https://www.annualreviews.org/doi/abs/10.1146/annurev-economics-042423-044154>). Dasgupta says: “Does the climate change literature assume, in effect, that it is the only sustainability problem? . . . [T]he established economics of climate change supposes that Nature provides us with only one maintenance and regulating service: carbon regulation. It permits analysts to imagine that with only modest investment in the transition to clean energy—say, 2% of global GDP until a net zero economy is attained—we can expect a future of indefinite GDP growth. . . . We do not have that luxury in the economics of biodiversity . . . The economics of climate change then becomes a branch of the economics of biodiversity (or of Nature writ large). So, instead of continually finding ways to reduce carbon concentration, as is customary in the economics of climate change, we search for ways to better manage our portfolio of assets in the economics of biodiversity. . . . It is not an accident that the bulk of the world’s biodiversity is in the tropics and that most of the world’s poorest people live there. Principal exports from those regions are primary products, whose extraction (from mines, plantations, wetlands, coastal waters, and forests) inflicts adverse externalities on local inhabitants. The externalities are not reflected in export prices, meaning that local ecosystems are overexploited; but that amounts to a transfer of wealth from the exporting country to the importing country, from a poor to a rich country. . . . Ecological externalities within regions suggest that countries in sub-Saharan Africa should collectively impose export taxes on primary products. That would ease pressure on their local ecosystems such as rainforests and fisheries and would also be a source of income for them.”

### **Follow-Up on Past JEP articles**

Gizelle George-Joseph and Devesh Kodnani of Goldman Sachs describe “Historically Black, Historically Underfunded: Investing in HBCUs” (Goldman Sachs Research, June 13, 2023, <https://www.goldmansachs.com/intelligence/pages/gs-research/the-case-for-investing-in-hbcus/report.pdf>). “[A]dmissions of Black students at HBCUs started to decline beginning in 1982, partly because historically white colleges had begun to admit Black students in greater numbers and Black colleges were unable to compete with the scholarships, facilities, or variety of academic programs offered by white colleges. The decrease in Black students was offset by the enrollment of non-Black students. According to the U.S. Department of Education Black students’ enrollment at HBCUs increased by 14% between 1976 and 2021; however, as the overall college enrollment of Black students more than doubled during that period, the share of Black students enrolled at HBCUs decreased from 18 percent in 1976 to just over 9 percent in 2021 . . . 59% of students at HBCUs receive Pell Grants compared to 31% at non-HBCUs. Moreover, 61% of HBCU students leverage Federal loans versus

40% of students at non-HBCUs . . . When some of these factors are controlled for, HBCUs have a higher graduation rate for Black college students than non-HBCUs.” George-Joseph and Kodnani emphasize and complement the article by Gregory N. Price and Angelino C. G. Viceisza in the Summer 2023 issue “What Can Historically Black Colleges and Universities Teach about Improving Higher Education Outcomes for Black Students?”

Jonathan Guryan and Jens Ludwig discuss “Overcoming Pandemic-Induced Learning Loss” (in *Building a More Resilient US Economy*, edited by Melissa S. Kearney, Justin Schardin, and Luke Pardue, Aspen Economic Strategy Institute, November 2023, pp. 150–170, <https://www.economicstrategygroup.org/publication/building-a-more-resilient-us-economy/>). “Because education is intrinsically cumulative, there is the real possibility that pandemic-induced school disruptions may set a whole generation of students off track for the rest of their lives . . . Students who can’t read at grade level by third grade are four times less likely to graduate high school. Ninth graders who have not yet passed their required entry-level math class (Algebra I) are five times less likely to graduate. . . . Overall, the learning gains from HDT [high-dosage tutoring] are much closer to offsetting the average learning loss experienced during the pandemic than other potential policy measures are. HDT is plausibly the intervention most up to the task of meeting the scale of our current learning-loss challenge. As one education expert put it, tutoring sessions are ‘the best learning conditions we can devise’ . . .” This essay complements “COVID-19, School Closures, and Outcomes,” by Rebecca Jack and Emily Oster, in the Fall 2023 issue.

Kristin Kiesel, Hairu Lang, and Richard J. Sexton discuss “A New Wave of Sugar-Sweetened Beverage Taxes: Are They Meeting Policy Goals and Can We Do Better?” (*Annual Review of Resource Economics*, 2023, pp. 407–432, <https://www.annualreviews.org/doi/abs/10.1146/annurev-resource-111522-111325>). “Indeed, it will take the knowledge and expertise of public health officials and scholars, economists, psychologists, and those most affected by health inequities and SSB [sugar-sweetened beverage] taxes to carefully design multifaceted policies that alter our food environments and nudge both producers and consumers toward improved behavioral responses, health outcomes, and greater social welfare. Carefully designed taxes on added sugars and unhealthy foods implemented countrywide could be part of combined policies aimed at reducing the obesity epidemic and related health harms. Given their regressivity, limited impact on consumption of SSBs, and failure to incentivize product reformulations, we find little basis to support further implementation of local SSB taxes.” This article updates the evidence and offers a differently nuanced conclusion than Hunt Allcott, Benjamin B. Lockwood, and Dmitry Taubinsky, “Should We Tax Sugar-Sweetened Beverages? An Overview of Theory and Evidence,” in the Summer 2019 issue of this journal.

## Discussion Starters

Thomas Peacock and Wendy S. Barclay argue that “Mink farming poses risks for future viral pandemics” (PNAS, July 19, 2023, <https://www.pnas.org/doi/10.1073/pnas.2303408120>). “Mink are highly susceptible to infection with several viruses that also infect humans. In late 2020, government agencies and academics in Europe and North America repeatedly documented that farmed mink had become infected with SARS-CoV-2, the causative agent of COVID-19. Evidence of mink-adapted viruses spilling back into local communities further demonstrated the poor biosecurity guidelines and practices in the industry. With this in mind, some countries—for example, the Netherlands—shut down mink production altogether. Fortunately, the mink-adapted variants of 2020 were not fitter than viruses circulating at the time in humans and, hence, did not spread widely. . . . [W]e argue that mink, more so than any other farmed species, pose a risk for the emergence of future disease outbreaks and the evolution of future pandemics.”

Nicholas Eberstadt and Ashton Verdery describe “China’s Revolution in Family Structure: A Huge Demographic Blind Spot with Surprises Ahead” (American Enterprise Institute, February 2023, <https://www.aei.org/research-products/report/chinas-revolution-in-family-structure-a-huge-demographic-blind-spot-with-surprises-ahead/>). “Our simulations show that the Chinese family is about to undergo a radical and historically unprecedented transition, as extended kinship networks atrophy across the nation and close blood relatives disappear altogether for many. This fraying of the extended family and atomization of the nuclear family come at an almost exquisitely inopportune moment in China: Social needs are soaring alongside the rising tally of elderly dependents and the shrinking ranks of those on whom the elderly can rely—two social indicators poised for inescapable collision in the years immediately ahead. Indeed, the withering of the Chinese family as we now know it will make for new and unfamiliar challenges at every stage in the life cycle, for both Chinese people and the Chinese state.”



## AEA INITIATIVES FOR DIVERSITY AND INCLUSION

For more details and information regarding how to apply for AEA diversity initiatives, please visit

[www.aeaweb.org/  
go/diversity-initiatives](http://www.aeaweb.org/go/diversity-initiatives)

**The American Economic Association is committed to the continued improvement of the professional climate in economics. In cooperation with key committees, the Association has launched several initiatives to support and promote diversity and inclusion in our profession.**

### 1 AEA Award for Outstanding Achievement in Diversity and Inclusion

This annual award will recognize departments and organizations that demonstrate outstanding achievement in diversity and inclusion practices. Focus will be on those applicants that take productive steps to establish new programs and procedures to create an inclusive environment, and to increase the participation of underrepresented racial/ethnic minorities, women, and LGBTQ+ individuals.

### 2 Departmental Seed Grants for Innovation in Diversity and Inclusion

These grants, in amounts up to \$5,000, will be awarded to economics departments to help establish new bridge programs or training programs for underrepresented minorities (URM). For example, a department might create a mentoring program for URM graduate or undergraduate students, create opportunities for URM students to do meaningful research assistant work, or start a program allowing URM students who need additional preparation for graduate school to take a lighter class load in the first year or to take core economics courses over two years.

### 3 The Andrew Brimmer Undergraduate Essay Prize

Thanks to the generosity of an anonymous donor, this paper prize has been established in honor of Andrew Brimmer, the first African American to serve on the Board of Governors of the Federal Reserve. The annual award will be presented to an undergraduate student at a US-based institution of higher learning majoring in economics, political science, public policy, or related fields for the best essay on the "economic well-being of Black Americans." The winner will receive a check for \$1,000 from the AEA.

### 4 URM Travel Grants

This award is open to junior economics faculty members from traditionally underrepresented groups in the economics profession. The grants will advance career and professional development by defraying the costs of travel, lodging, and conference registration to attend the annual ASSA Meeting.

### 5 Small Group Breakfast Meeting for URM Scholars

Each year at the ASSA Meeting, a breakfast with prominent economists is held for underrepresented minority scholars. The goal is to allow URM scholars access to AEA journal editors, executive board members, thought leaders in specific areas of economics, or other economists for the purpose of addressing issues of access to journals, conferences, and networks that are often out of reach for URM scholars.

### 6 Professional Development Grant for URM Faculty

This \$2,000 grant was established to help advance the career and professional development of underrepresented minority faculty members in the field of Economics. The award is open to eligible junior economics faculty members. Entrants to the essay competition should detail their research and how it relates to economics education.

*These initiatives are another important step in making our field accessible and welcoming to anyone with the interest and ability to pursue a career in economics. Please help us share this information throughout the profession so we can all work together and continue to improve.*

# JOE NETWORK

The career network designed by economists—for economists.

**The AEA's JOE Network is the preferred hiring tool for the economics job market, matching qualified candidates with relevant economics positions in leading institutions.**



## Job Seekers

Showcase your profile to hundreds of hiring managers and quickly apply to your preferred positions. The JOE Network allows you to securely request reference letters and share job market materials.

## Employers

Target search requirements to locate your preferred candidates, set up hiring committee access, and easily manage job postings and applications.

## Faculty Letter-Writers

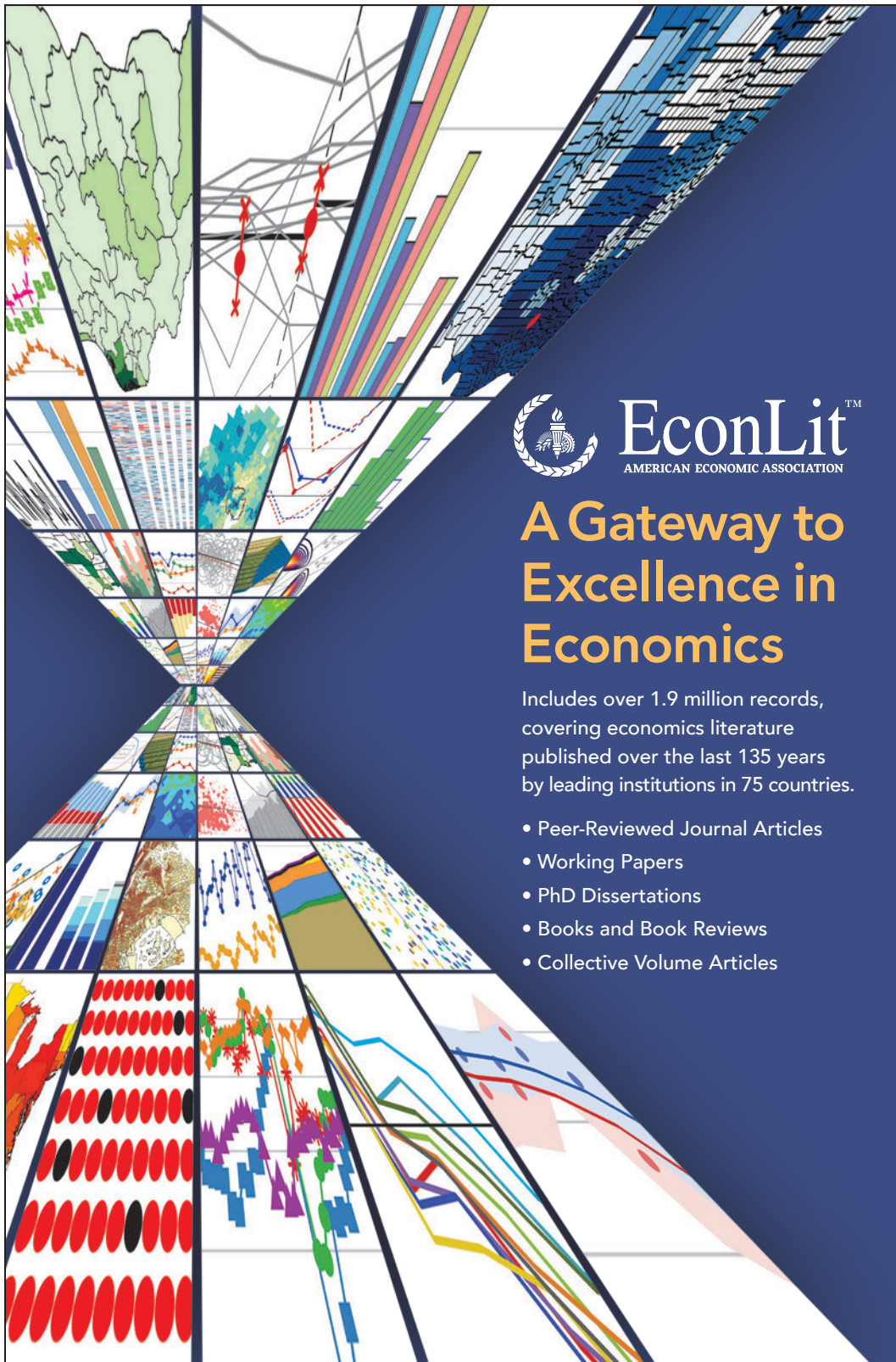
Reply to reference requests with default or customized letters, assign proxies, and easily monitor task completion. Over 150,000 requests are fulfilled on the JOE Network each year.

Use the JOE Network as your comprehensive tool for the economics job market!



[aeaweb.org/JOE](http://aeaweb.org/JOE)





**EconLit**<sup>TM</sup>  
AMERICAN ECONOMIC ASSOCIATION

## A Gateway to Excellence in Economics

Includes over 1.9 million records,  
covering economics literature  
published over the last 135 years  
by leading institutions in 75 countries.

- Peer-Reviewed Journal Articles
- Working Papers
- PhD Dissertations
- Books and Book Reviews
- Collective Volume Articles

Find more information at [www.econlit.org](http://www.econlit.org) or contact your economics professor  
or university librarian about gaining access to *EconLit*.



# Join Us in Shaping the Future of Economics

## Become an AEA Member!

As a member, your partnership with the AEA provides you with distinct career advantages and enables us to elevate the profession for all economists.

- ▶ Build lifelong relationships with mentors and peers
- ▶ Access essential research and job resources
- ▶ Publish and present your research to colleagues
- ▶ Gain meaningful support and advice on your work
- ▶ Collaborate on solutions to issues you care about

Today our community consists of 18,000+ individuals who are committed to the advancement of economics and its enduring contributions to society. We welcome the opportunity to help you advance your career in economics.

**Aim High. Achieve More. Make a Difference.**



AEA Journals



@AEAJournals / @AEAInformation



Join today!

[aeweb.org/membership](https://aeweb.org/membership)



# 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](mailto: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>. Change of address notice must be received at least six weeks prior to the publication month.

Copyright © 2024 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 the 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](mailto:aeainfo@vanderbilt.edu).

## EXECUTIVE COMMITTEE

### Elected Officers and Members

#### *President*

JANET CURRIE, Princeton University

#### *President-elect*

LAWRENCE F. KATZ, Harvard University

#### *Vice Presidents*

NICHOLAS BLOOM, Stanford University

EBONYA L. WASHINGTON, Columbia University

#### *Members*

AMANDA BAYER, Swarthmore College

MELVIN STEPHENS, JR., University of Michigan

KEVIN LANG, Boston University

LISA M. LYNCH, Brandeis University

AMANDA E. KOWALSKI, University of Michigan

LARA D. SHORE-SHEPPARD, Williams College

#### *Ex Officio Members*

CHRISTINA ROMER, University of California, Berkeley

SUSAN C. ATHEY, Stanford University

### Appointed Members

*Editor, The American Economic Review*

ERZO F.P. LUTTMER, Dartmouth College

*Editor, The American Economic Review: Insights*

MATTHEW GENTZKOW, Stanford University

*Editor, The Journal of Economic Literature*

DAVID H. ROMER, University of California, Berkeley

*Editor, The Journal of Economic Perspectives*

HEIDI WILLIAMS, Dartmouth College

*Editor, American Economic Journal: Applied Economics*

BENJAMIN OLKEN, Massachusetts Institute of Technology

*Editor, American Economic Journal: Economic Policy*

LUCAS W. DAVIS, University of California, Berkeley

*Editor, American Economic Journal: Macroeconomics*

AYŞEGÜL ŞAHİN, University of Texas at Austin

*Editor, American Economic Journal: Microeconomics*

NAVIN KARTIK, Columbia 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

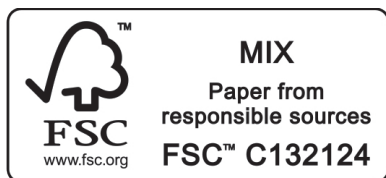
### ADMINISTRATORS

*Director of Finance and Administration*

BARBARA H. FISER

*Convention Manager*

REBEKAH LOFTIS



*The Journal of*  
***Economic Perspectives***

---

Winter 2024, Volume 38, Number 1

---

**Symposia**

***The Global Market for Talent***

**Britta Glennon,**

“Skilled Immigrants, Firms, and the Global Geography of Innovation”

**Francesco Lissoni and Ernest Miguelez,**

“Migration and Innovation: Learning from Patent and Inventor Data”

***Taxation and Developing Countries***

**Pierre Bachas © Anders Jensen © Lucie Gadenne,**

“Tax Equity in Low- and Middle-Income Countries”

**Oyebola Okunogbe and Gabriel Tourek,**

“How Can Lower-Income Countries Collect More Taxes?  
The Role of Technology, Tax Agents, and Politics”

**Anne Brockmeyer © Giulia Mascagni © Vedanth Nair**

**© Mazhar Waseem © Miguel Almunia,**

“Does the Value-Added Tax Add Value? Lessons Using Administrative  
Data from a Diverse Set of Countries”

**Articles**

**Andrew Metrick,** “The Failure of Silicon Valley Bank and the Panic of 2023”

**Zoë Cullen,** “Is Pay Transparency Good?”

**Olivier Marie and Paolo Pinotti,**

“Immigration and Crime: An International Perspective”

**Nancy Folbre,** “Care Provision and the Boundaries of Production”

**Eliana Carranza and David McKenzie,** “Job Training and  
Job Search Assistance Policies in Developing Countries”

**Features**

**Timothy Taylor,** “Recommendations for Further Reading”

