The Journal of

Economic Perspectives

A journal of the
American Economic Association

Spring 2021

The Journal of
Economic Perspectives

A journal of the American Economic Association

The Journal of
Economic Perspectives

Contents   *Volume 35 • Number 2 • Spring 2021*

Symposia

Statement of Purpose

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

Policy on Data Availability

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

Policy on Disclosure

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

The Resilience of the Euro

Philip R. Lane

The euro launched in 1999, with the euro area initially consisting of eleven member countries of the European Union. The replacement of individual national currencies with a single transnational currency was a significant innovation in international monetary economics and involved the establishment of the European Central Bank to set the monetary policy and operate the new monetary system (together with the national central banks, collectively known as the Eurosystem) for the euro area.

As a monetary system, the euro area looked to have worked reasonably well in its first decade—perhaps better than expected, given the magnitude of the transition. But the euro area then experienced a prolonged crisis during the 2007–2013 period, which raised many questions about its resilience. It seems like a good time to revisit the experience of the euro area: in particular, the post-crisis record over 2014–2019 and the response to the pandemic shock in 2020 provide important new evidence in examining the resilience of the single European currency.

A standard approach for economists analyzing the euro has been to focus on the "optimum currency area" literature pioneered by Mundell (1961): for discussion, see O'Rourke and Taylor (2013 in this journal) and Corsetti et al. (2020). If shocks are mainly common and symmetric among the member countries, a common

■ *Philip R. Lane is Member of the Executive Board and Chief Economist, European Central Bank, Frankfurt, Germany. He is also Honorary Professor of Economics at Trinity College Dublin, Ireland, and Research Fellow, Centre for Economic Policy Research, London, United Kingdom.*

monetary policy may be collectively preferable to setting national monetary policies, especially if there is a high degree of economic and financial interdependence among the member countries. For instance, in the context of the highly open EU economy, a common currency might provide valuable insulation from externally driven fluctuations in the world economy and the global financial system.

Alternatively, if asymmetric shocks among the member countries are prevalent, a geographic area will be better suited to have a single currency (and will not require the flexibility of exchange rates between national currencies) if it has other internal mechanisms for facilitating adjustments between countries, including well-integrated product and labor markets, a fully integrated financial system, and a substantial federal fiscal system. In providing a benchmark for these criteria, this line of argument often compares the euro area to the United States and identifies a significant shortfall for the euro area. However, such an assessment cannot be static: after all, the US economy did not fit the optimum currency criteria especially well for the first 100–150 years of its history. Similarly, the institutional framework for the euro area has been reworked in recent years, precisely in reaction to the emergence of costly gaps in the original institutional framework of the euro area.

In any event, a realistic evaluation of the coherence of the euro area as a monetary union should not only refer to the optimum currency area criteria as a checklist, as if the euro was concocted in an academic seminar by a group of technocrats. Rather, the creation of the euro and the ongoing evolution of euro area institutions have been driven by political choices. Any real-world assessment must be weighed in the context of the political economy underpinning the euro and the European Union.

This paper will discuss some central issues in understanding the resilience of the euro area. First, rather than only comparing the euro area to unobservable counterfactual scenarios (how Europe might have fared with the retention of national currencies), a basic criterion is whether the euro delivers reasonable macroeconomic outcomes. After all, member states would find it hard to stand behind the single currency if it manifestly failed to deliver macro-financial stability and/or was universally dominated by alternative monetary regimes. In assessing macroeconomic outcomes, it is necessary to take account of the differences across the three phases of the euro's history, which can be broadly characterized as the initial years of 1999–2007; the extended crisis period of 2008–2013; and the recovery phase of 2014–2019. The onset of the pandemic in 2020 in effect constitutes a fourth phase in the evolution of the euro area.

In relation to the recovery phase since 2014, the costly lessons of the crisis period (including the retrospective diagnosis of the shocks and policy errors that contributed to pre-crisis imbalances) may have contributed to better national and EU-level policymaking in the post-crisis period. In particular, there was a significant stabilization of macro-financial conditions over 2014–2019, albeit with meaningful differences across the member countries. In understanding the post-crisis stability of the euro area, it is also essential to acknowledge that, as a result of the global

trend decline in the equilibrium real interest rate, the high outstanding (public and private) debt stocks in some European countries have been much less salient than feared, because low interest rates have kept debt servicing burdens lower than might have been expected.[1]

Second, while macroeconomic performance did generally improve in the euro area over 2014–2019, this was a relatively benign period in terms of global economic and financial conditions. Accordingly, the current pandemic crisis provides a more severe test of the resilience of the euro. In particular, I will discuss the reforms of the euro area institutional architecture launched in 2012. The pandemic shock represents an important initial test of whether these institutional reforms have improved the stability and sustainability of the euro area.

Third, I revisit the political economy foundations of the euro. It is sometimes underappreciated that the euro is closely intertwined with the broader institutional framework of the European Union: the deep commitment by the member states to the European Union is an extremely strong source of political backing for the euro (Whelan 2019). Indeed, much of the debate concerning the resilience of the euro ultimately turns on the status of the European Union as a shared institutional commitment among the member countries (for an excellent guide to the political economy of the European Union in this journal, see Spolaore 2013). The intertwining of the euro area and the European Union is even stronger now that the United Kingdom has left the European Union in the "Brexit" process: the 19-member euro area now constitutes 85.4 percent of the GDP of the 27-member European Union compared to 71.6 percent of the GDP of the 28-member pre-Brexit EU composition.

In overall terms, my assessment is that there has been an increase in the relative importance of common shocks versus asymmetric shocks in the EU economy in recent years, which has improved the relative attractiveness of a common currency and has acted to strengthen the coherence and stability of the euro area. While the resolution of the 2008–2013 crisis was very costly (especially in the most indebted countries), the set of post-crisis reforms to the euro area institutional architecture means that it is less likely that such destabilizing dynamics can take hold in the future. Taken together, the increase in the relative importance of common shocks, the strengthened capacity to manage asymmetries among the member countries, and the permanent introduction of crisis management tools mean that the euro is likely to prove far more resilient than was predicted by many commentators during the 1990s debate on the formation of the monetary union and during the 2010–2012 debate on the capacity of the euro area to overcome the forbidding challenges it faced at that time.

---

[1] The trend decline in the equilibrium real rate has also transformed central banking, with the European Central Bank joining other central banks in using large-scale asset purchases as a monetary policy tool, in addition to setting the short-term policy rate.

Macro-Financial Outcomes in the Euro Area

As noted above, the first 20 years of the euro should be sub-divided into three broad phases: the pre-crisis period (1999–2007); the twin crises (global financial crisis and euro sovereign debt crisis) period (2008–2013); and the recovery period (2014–2019).[2] Of course, the pandemic shock of 2020 constitutes a recent stark break in the data.

The pre-crisis and crisis phases have been much studied (Lane 2006, 2012, 2019b; Bastasin 2012; Sandbu 2015; Brunnermeier, James, and Landau 2016; Schelkle 2017; Mody 2018).[3] While my primary focus in this article is on the post-crisis recovery period, it is perhaps useful to provide a quick recap of the 2010–2012 euro sovereign debt crisis that sharply differentiated the euro area from other advanced economies. In particular, many elements of the pre-2010 period were shared across all advanced economies, with the structure of the world economy reshaped by the twin forces of globalization and technological advances, while there was a significant easing in international financial conditions (especially during 2003–2007). In turn, this fostered significant construction booms and surges in property prices in some countries (and an expansion in fiscal deficits in some other countries), which were associated with an amplification of current account imbalances and funded by a surge in cross-border lending. The dynamics underlying these imbalances came to a painful sudden stop during the 2008–2009 global financial crisis.

In terms of the contribution of the euro to the pre-crisis imbalances, the launch of the euro in and of itself constituted a major asymmetric shock, in view of the very different initial positions of the member countries. In particular, adopting the euro meant a sharp drop in nominal interest rates in the lower-income member countries, with attendant implications for asset prices and credit markets. In combination with the global forces listed above and the lack of sufficient countervailing policy responses at national and EU levels, this euro-specific shock contributed to the increase in the magnitude and persistence of imbalances in the euro area in the run up to the global financial crisis.

Most fundamentally, the euro area lacked collective crisis management tools in relation to banking crises or fiscal crises: this deficiency was a major factor in converting the 2008–2009 global financial crisis into the euro area sovereign debt crisis during 2010–2012. I return to the topics of crisis prevention and crisis management later; for now, I shift forward to focus on macroeconomic performance during the 2014–2019 post-crisis period.

Figure 1 shows the distribution across the member countries of the main macroeconomic variables that dominate the policy debate: output, inflation, the current

---

[2] These timeframes should be regarded as schematic; for example, the initial phases of the financial crisis can be traced to August 2007; similarly, although 2014–2019 was a period of improved performance, it did include some major risk episodes, most prominently the tensions over an official funding program for Greece in spring 2015.

[3] Comprehensive analyses of the first two decades of the euro from a monetary policy perspective are provided by Hartmann and Smets (2018) and Rostagno et al (2019).

Figure 1
Macroeconomic Indicators in the Euro Area
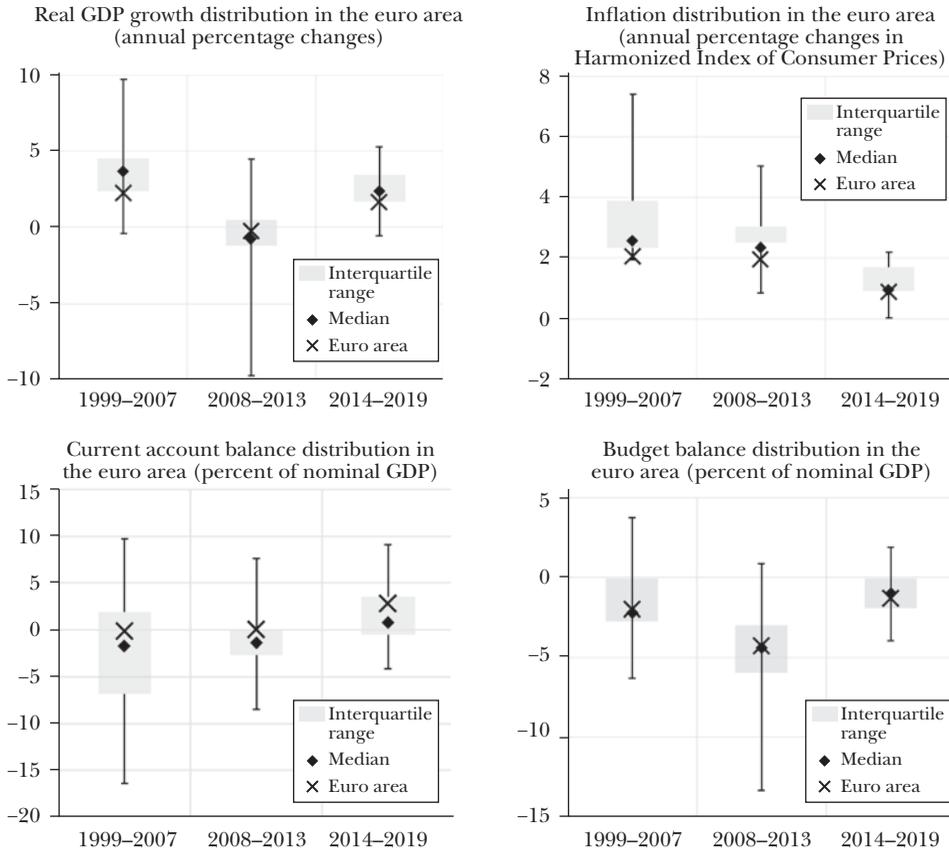
account balance and the fiscal balance. Whereas output and inflation capture aggregate macroeconomic performance in many models, the current account balance and the fiscal balance are also tracked in order to assess the underlying sustainability of macroeconomic performance.

The upper left panel shows a sustained output recovery over 2014–2019, which is also reflected in the decline in the euro area unemployment rate from its peak annual rate of 12.0 percent in 2013 to 7.5 percent in 2019. In relation to the upper right panel, there had been a gradual pickup in inflation in recent years relative to

the very low post-crisis levels, even if inflation has remained below the inflation aim of the European Central Bank of "below, but close to, two per cent."

Across the range of indicators, the clear pattern is that cross-country dispersion has been much lower in the post-crisis period compared to the pre-crisis period (in capturing dispersion across the member countries, my primary focus is on the inter-quartile range, because the extreme values are distorted by well-documented quirks in the macroeconomic accounts of small countries). In particular, the incidence of very poor macroeconomic outcomes has declined and the risk-amplifying configuration of high fiscal deficits and high current account deficits was virtually eliminated during this period. These changes went a long way to defusing the widespread concerns about the future of the euro that were prevalent during the worst of the crisis period. While the resilience of a monetary union is enhanced by less cross-country heterogeneity in macroeconomic dynamics and lower macroeconomic tail risk, the rebalancing episodes were predominantly asymmetric, mainly taking the form of expenditure reduction in the excessive-deficit countries rather than expenditure expansion in the excessive-surplus countries. The overall patterns during the post-crisis period of a persistent aggregate current account surplus and a marked reduction in the aggregate fiscal deficit also reflected the absence of a coordinated or joint approach to ensuring that the area-wide fiscal stance was aligned with overall macroeconomic conditions, given that monetary policy was the only policy instrument calibrated for the euro area as a whole.

Even if the "flow" imbalances—the current account imbalance and the fiscal imbalance—had become less negative, the "stock" imbalances embedded in balance sheets also represent sources of macro-financial risk. Figure 2 plots the stocks of public debt, private-sector debt, and net external debt, together with the distribution of capital-asset ratios for the banking system. Debt stocks rose substantially during the crisis, especially public debt, and remained quite high at the end of 2019. However, there was some gradual general decline in debt ratios since the earlier euro crisis. Moreover, the vulnerabilities embedded in high debt stocks have been partly mitigated by the sustained increase in capital-asset ratios in the banking system in the post-crisis period, shown in the lower right panel of the figure. This has improved the loss-absorbing capacity of the banking system, which was so lacking during the 2008–2013 period. Still, the remaining high dispersion in debt stocks represented a significant source of heterogeneity across the euro area.

As it turned out, the pervasive post-crisis concerns about the sustainability of high legacy debt levels were substantially attenuated by the remarkable shift in interest rates, as shown in Figure 3. Very low debt servicing costs did much to improve the dynamics of highly-indebted countries and sectors. In addition, sovereign risk premia have narrowed substantially, even if there was a temporary spike during the initial stages of the current pandemic shock. Very low risk-free rates and low risk premia mean that the differences in outstanding debt stocks across the member countries do not map into substantial differences in debt servicing burdens and also limit the self-feeding impact of high interest rates on debt dynamics.

Figure 2
Balance Sheet Indicators for the Euro Area

Government debt
(percent of nominal GDP)

Private sector debt
(percent of nominal GDP)

Net external debt
(percent of nominal GDP)

Capital adequacy ratio
(percentages)

Interquartile range ◆ Median
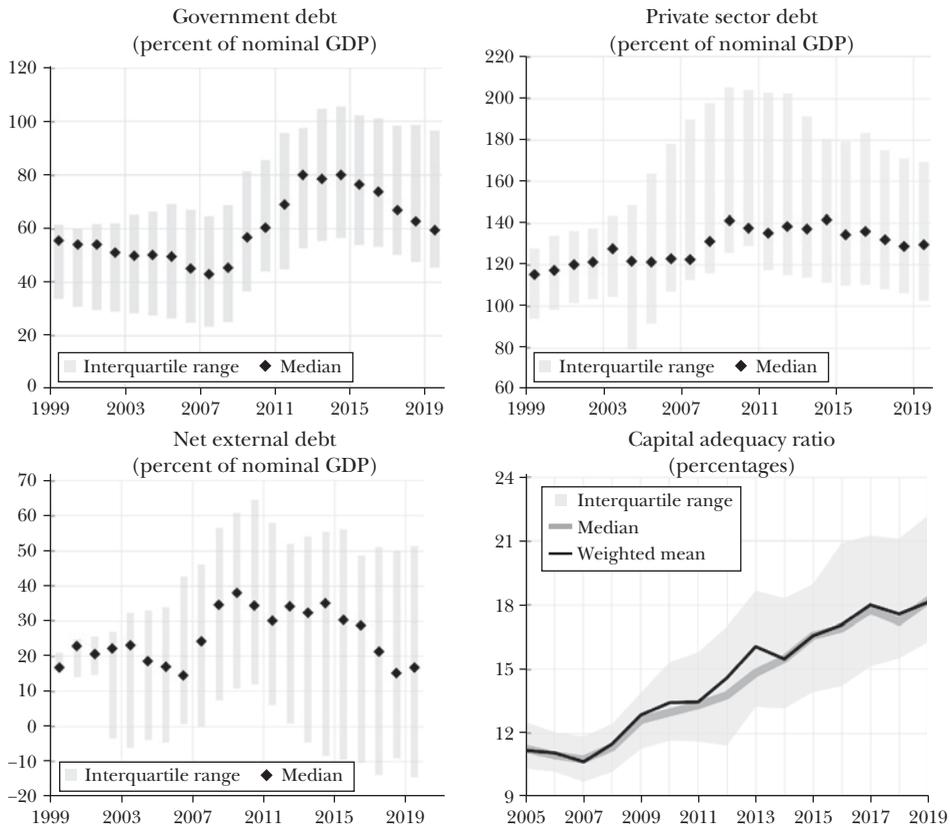
Note: Panel B: Private *se*ctor gross debt refers to the consolidated definition and is the sum of NFC and household consolidated gross debt. NFC consolidated gross debt is defined as the sum of total loans granted to NFCs net of intra-sectoral lending, debt securities issued and pension liabilities. Household consolidated gross debt includes total loans granted to households net of intra-sectoral lending. Panel C: The interquartile ranges are calculated based on incomplete country data. Data are only available as of 2000 for Latvia, 2002 for Ireland and Luxembourg, 2003 for Greece and the Netherlands, 2004 for Lithuania, Slovakia and Malta, 2005 for Belgium and, 2008 for Cyprus and France. For all other countries data are available as of 1999. Euro area net external debt prior to 2008 is proxied as the sum of net liabilities of direct investment debt instruments, portfolio investment debt instruments and other investment minus reserve assets excluding monetary gold. Panel D: The capital adequacy ratio is computed as total capital over risk-weighted assets. It is based on an unbalanced sample of banks located in one of the 19 euro area countries, irrespective of the time of adoption of the euro. The sample encompasses 50 entities in 2005 and 113 in 2019.
Source: Upper left panel: European Commission and European Central Bank calculations. Upper right panel: Eurostat and European Central Bank calculations. Lower left panel: European Central Bank (Balance of Payments) and Eurostat. Lower right panel: S&P Market Intelligence (SNL Financial) and European Central Bank calculations.

Figure 3
Ten-year Euro Area Bond Yields: GDP-weighted and Selected Countries
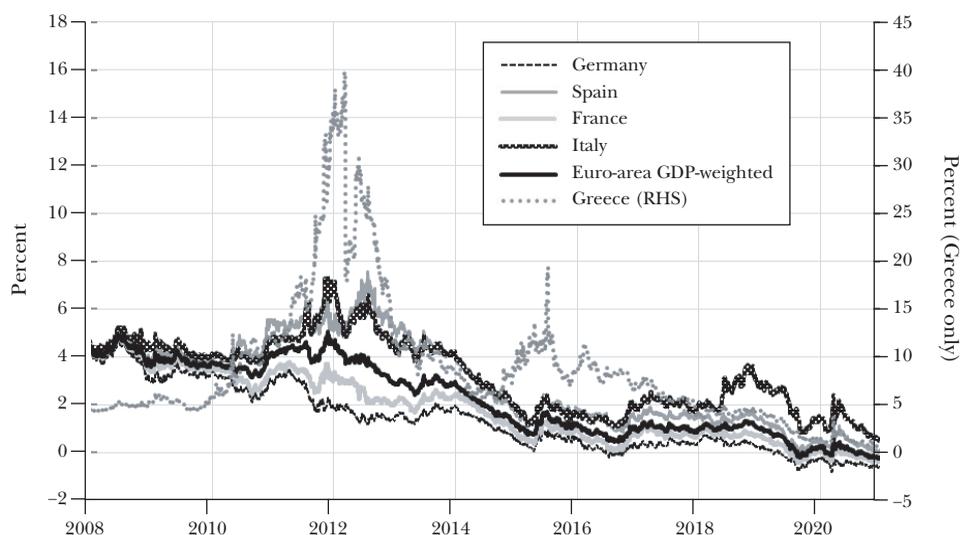
Source: Refinitiv, Bloomberg, and European Central Bank calculations. Latest observation: December 31, 2020.

In large part, this radical shift in the interest rate environment reflects the worldwide decline in the underlying equilibrium real interest rate that can be linked to trends in demographic, productivity, and risk preferences (Lane 2019c). At a cyclical level, it also reflects the accommodative monetary policy stance since 2014 that has been required by the persistence of below-target inflation.

In the context of the euro area, the compression in sovereign yields can also be attributed to a reassessment of the risks associated with holding euro-denominated sovereign bonds. During 2010–2012 especially, the precedent of the Greek sovereign debt restructuring and the severe imbalances in some member countries had prompted a sharp increase in sovereign risk premia, with global investors pricing in that sovereign debt restructuring was a possible outcome for euro area member states. Furthermore, default risk was compounded by the lack of euro area crisis management tools and unanswered questions about the collective and country-by-country commitment to the integrity of the monetary union under all scenarios.

The accumulated track record during the relative calm of the 2014–2019 period served to reduce risk premia by providing some evidence that the imbalances that emerged in the first decade of the euro were not intrinsic to the design of the monetary union. Instead, better policies at national and European levels helped to deliver improved macro-financial stability for the euro area and the individual member countries. In particular, 2014–2019 might be viewed as a more typical period in

terms of what should be expected on an ongoing basis: under this reading, the initial years of the euro could be interpreted as a one-off "learning" phase, with a too-slow policy adjustment to the realities of living with a single currency.[4]

The completion of the EU-IMF adjustment programs, in the member countries that ultimately had to seek official-sector funding support, provided important evidence of the willingness of member countries to undertake costly adjustments rather than turn to sovereign default or exit from the euro as methods to resolve excessive imbalances. Greece (2010–2018), Ireland (2010–2013), Portugal (2011–2014), and Cyprus (2013–2016) all undertook adjustment programs, financed by an official funding mix of EU and IMF sources. In addition to significant fiscal corrections and the restructuring of the banking system in all of these cases, both Cyprus and Greece introduced capital controls, and Cyprus imposed haircuts on bank depositors (in excess of €100,000). In addition, Spain received EU official funding to finance the restructuring of its banking system in 2012–2013. In all cases, the official funders continue to monitor these economies on an enhanced basis, because the payback period for the official funds extends into the distant future. This experience indicated that all member states were highly committed to the euro and ultimately willing to take tough measures in order to ensure compatibility with euro area membership. While the greatest test was for those countries that had to implement difficult adjustment measures, the willingness of the other member countries to back official loans to this group was also significant, highlighting both the high potential spillovers inside a monetary union and the significant political interdependence across all the member countries.

In addition, a fundamental source of euro resilience is that it remains a forbidding prospect for any stressed country to consider exiting the euro area, in view of the substantial disruption likely to be associated with such a move (Eichengreen 2010). This also holds true for the general membership of the euro area: if any individual country were to seek to exit the euro, there would be a considerably heightened risk of contagion forces and even of a collective break-up of the monetary union.

This recovery phase for the European economy from 2014–2019 was also supported by the steps taken to improve the institutional design of the euro area (including measures to reduce the likelihood of future crises and improve crisis management tools), which will be discussed in the next section. The foundations of the political commitment to the euro are analyzed in the following section.

The Institutional Architecture of the Euro Area

The 2008–2013 twin crises in the euro area induced a series of reforms of the institutional setup of the euro area (for an analytical overview of main elements

---

[4]Accounts of the policy errors made in the first decade of the euro include Shambaugh (2012), Fernández-Villaverde, Garicano, and Santos (in this journal, 2013), and Honohan (2019).

of the euro crisis, see Brunnermeier and Reis 2019). During this period, the most urgent task was to develop an effective crisis management framework. But it was also essential to reduce the likelihood of future crises by implementing crisis prevention measures, improving the resilience of the euro area, and establishing safeguards against the re-emergence of the macro-financial imbalances that proved to be the key drivers of the twin crises.

Crisis Prevention Measures

In relation to crisis management, the initial design of the euro area architecture had not envisaged any need for non-market funding of member states. Implicit in this setup was that, in the unlikely event that a member country needed to restructure its debt, it would do so on its own. But in spring 2010, the Greek government was no longer able to tap private capital markets. In this situation, the case for official funding (as opposed to just relying on sovereign debt restructuring) in the euro area was broadly the same as for other countries (typically emerging or developing countries) facing a funding crisis: a temporary phase of non-market funding could provide the time to execute a macro-financial adjustment program that could restore market confidence. Indeed, it was conjectured that contagion risk might be even more severe inside a monetary union, which further tilted the argument in favor of official funding.

These factors led to the European Union designing an ad hoc official funding rescue package (in combination with the International Monetary Fund). Ultimately, the scale of the Greek fiscal imbalance was so severe that an extensive sovereign debt restructuring also proved necessary in spring 2012.

By 2012, it was clear that a systemic approach to providing official funding would be necessary to safeguard financial stability, and a permanent mechanism called the European Stability Mechanism (ESM) was established. By providing certainty that official funding would be available under specified conditions if needed, the very existence of the ESM calmed markets. By providing a backstop funding source, it deterred speculative attacks and thereby reduced the risk to private investors of rolling over sovereign debt.[5] The euro area member countries committed paid-in capital of €80 billion to the ESM: this equity allows it to issue highly rated bonds, with the proceeds lent to those countries requiring official assistance. The total lending capacity of the ESM was set at €500 billion.

There was one way in which the European Stability Mechanism deviated from standard doctrine. The usual rule is that an official lender-of-last-resort should set a penalty interest rate in order to deter excessive use of the facility, but the ESM ultimately decided to provide low-cost, long-duration official loans (conditional on adopting policies that provided a pathway to more stable public finances). The decision to set a low interest rate and a long horizon for the repayment of ESM loans

---

[5]As analyzed by Bianchi and Mondragon (2019), it might be argued that rollover risk is higher inside a monetary union, in view of the tougher adjustment path in the absence of a national currency devaluation option (see also Corsetti et al. 2014).

acted to reduce the present value of the debt owed by program countries, which significantly alleviated the adjustment challenges they faced.

In turn, with solvency concerns addressed by the availability of conditional official loans from the European Stability Mechanism, the European Central Bank could then commit to ensuring liquidity in the sovereign debt market for program countries. In July 2012, the President of the European Central Bank, Mario Draghi, made his famous "whatever it takes" declaration. The speech had an immediate calming impact in sovereign debt markets, since investors took reassurance that the European Central Bank would not permit a self-fulfilling "bad" equilibrium by which a loss of investor confidence could trigger default by a solvent-but-illiquid sovereign. Soon after, the European Central Bank formalized this commitment with its Outright Monetary Transactions (OMT) program.

The European Central Bank has never had to activate the OMT program: it has primarily worked by eliminating self-fulfilling liquidity runs in euro area sovereign debt markets. However, the "double protection" provided by the ESM and the OMT programs were decisive in calling a halt to the intense phase of the euro area sovereign debt crisis in autumn 2012.

Improving the Resilience of the Euro

Turning to the prevention of future crises, the global financial crisis and the euro crisis from 2008–2013 taught (at least) three crucial lessons.

First, macroprudential policy should play a significant role in preventing imbalances and safeguarding financial stability. By limiting the leverage of households, firms, and banks through a mix of borrower-based and lender-based restrictions, boom-bust dynamics are muted, making it is less likely that large imbalances can accumulate and improving the capacity to absorb adverse shocks (useful starting points include Farhi and Werning 2016; Martin and Philippon 2017).[6] The potential value of macroprudential policy has become mainstream in the global macro-financial policy community and is part of the "integrated policy framework" currently under development at the IMF.

Second, the resilience of the banking system should be improved by a combination of increasing capital-asset ratios and new laws governing the resolution of failing banks, complemented by a more intrusive approach to banking supervision. Several steps were taken along these lines. In line with the global adoption of higher capital buffers for banking systems, the capital asset ratio of euro area banks climbed in the post-crisis period (as shown in Figure 2 earlier). The European Union established the Single Supervisory Mechanism by which the European Central Bank would become responsible for area-wide banking supervision. This serves to distance bank supervisors from national pressures and establish an area-wide level playing field. Most importantly, a common set of banking regulations is a key step in establishing the

---

[6]The pan-European monitoring of systemic risks and incipient macro-financial imbalances was also enhanced by the creation of the European Systemic Risk Board (ESRB) and the introduction of the Macroeconomic Imbalance Procedure (MIP) system.

conditions under which an area-wide banking union could properly develop. A full banking union would act to insulate the banking system from national fiscal shocks, in turn attenuating the fiscal-bank "doom loop" that was so damaging during 2008–2013.

The bank reform agenda also included additional steps with their own euro-acronyms. A Bank Recovery and Resolution Directive (BRRD) made it easier to resolve or shut down failing banks. An area-wide Single Resolution Mechanism (SRM) was established, with a new Single Resolution Board (SRB) charged with acting as the area-wide central resolution authority, and banks across the euro area making contributions to the Single Resolution Fund (SRF) that can be deployed as a funding source for the resolution of systemically-important banks. The increased coordination and centralization of financial supervision was further supported by the creation of the EU supervisory agencies: the European Banking Authority (EBA), European Securities and Markets Authority (ESMA), and the European Insurance and Occupational Pensions Authority (EIOPA).

One item still lacking is area-wide protection of small depositors through a common European Deposit Insurance System (EDIS). Risk exposures significantly declined during 2014–2019 through the reduction in non-performing loans, the increase in capital buffers in banking systems, the implementation of macro-prudential policy frameworks, and the stabilization of fiscal ratios and improved macroeconomic performance (for discussion, see Carmassi et al. 2018). However, the balancing between risk reduction and risk sharing in determining the transition to European-wide deposit insurance remains a live policy debate. The remaining reform agenda has also included efforts to develop larger and more integrated markets for equity and debt securities (capital markets union), which would improve the resilience of the financial system by reducing dependence on banks and facilitating pan-European risk sharing, including sharing the risks embedded in the financing of the banking system.

The third element in the crisis prevention framework was to improve the conduct of national fiscal policies (for additional discussion, see the essay by Bilbiie, Monacelli, and Perotti in this symposium). The crisis-induced surge in public debt ratios illustrated the value of fiscal buffers in managing tail shocks and conversely showed how excessively high debt reduced resilience. In order to add to the technical quality of fiscal debates, the plan was that national fiscal councils would issue independent opinions on the annual budget plans, and EU-level probing of national fiscal plans would be stepped up by the European Commission and through the advisory European Fiscal Board (EFB).

To underpin fiscal sustainability, there was a concentrated effort from 2010 onwards to unwind the large deficits that had emerged during the global financial crisis. There is by now broad agreement that the collective post-crisis pace of fiscal correction across the European Union was too severe in terms of its macroeconomic impact. Nonetheless, a reset of the European fiscal framework seemed to be a political precondition for the other measures (like the European Stability Mechanism) that required an increase in joint contingent fiscal liabilities. Fiscal stabilization also facilitated the initiation of quantitative easing by the European

Central Bank, which has been a significant purchaser of national sovereign bonds since 2015. More recently, the role of national fiscal policies in area-wide macroeconomic stabilization has been increasingly acknowledged, with the European fiscal framework implemented in a flexible manner to avoid clashes between the fiscal rules and macroeconomic objectives.

In addition to safeguarding the sustainability of national fiscal positions, the post-crisis EU reform agenda has also included a debate on deepening the extent of fiscal union by introducing a "central fiscal capacity" to foster area-wide fiscal risk sharing and reflect the area-wide macroeconomic situation. In addition, a deeper fiscal union would also copperfasten the EU banking union by providing a truly common fiscal backstop for the financial system (Marzinotto, Sapir, and Wolff 2011). A larger EU-level budget could also be an efficient approach to funding shared public goods.

It is not straightforward to design a fiscal union that balances European-wide risk sharing while preserving incentives for prudence in national policy decisions (Farhi and Werning 2017; Bénassy-Quéré et al. 2018). Among other reasons, it is not straightforward to analyze the relative contributions of "bad luck" versus "bad policy choices" in driving asymmetric budgetary outcomes across countries. In addition, simulations indicate that the size of a risk-sharing program would have to be quite large in order to have a macroeconomic impact (for examples, see Arnold et al. 2018; Berger, Dell'Ariccia, and Obstfeld 2018).

The coronavirus crisis of 2020 provides a test case on whether a monetary union can be resilient with only national fiscal policies (as also discussed by Bilbiie et al. in this symposium). The initial EU-level fiscal response in April 2020 had three elements. First, the European Union would collectively borrow up to €100 billion (about 0.7 percent of EU GDP) to provide low-cost loans to national governments to support employment-preservation policies, under the new SURE initiative (which stands for Support to mitigate Unemployment Rising in an Emergency). Second, the European Stability Mechanism (ESM) would provide contingent credit lines up to a ceiling of €240 billion (about 1.7 percent of EU GDP) to member states, with no country allowed to draw more than 2 percent of its own GDP. Third, the member states would provide guarantees of €25 billion to enable the European Investment Bank (the EU public bank) to scale up lending to small and medium enterprises by €200 billion (about 1.4 percent of EU GDP).

While these EU-level initiatives were welcome, they were relatively small in scale and were essentially designed as mechanisms to support national fiscal actions. It was soon evident that a more ambitious EU-level fiscal response was needed.

A State-Contingent Fiscal Union?

In July 2020, a new joint fiscal initiative was agreed upon: the NGEU (Next Generation EU) recovery instrument. This is quantitatively and qualitatively different than the other programs (for details on the NGEU, see https://www.consilium.europa.eu/media/45109/210720-euco-final-conclusions-en.pdf).   At a total size of €750 billion, it corresponds to about 5.3 percent of EU GDP. It is

divided roughly in half, with €360 billion for low-cost loans to national governments (in a similar vein to the SURE program to support employment preservation) and €390 billion allocated to EU grants to support a range of spending programs. These grants are equivalent to federal-level spending programs in the US system, with no direct connection between the level of spending in a given state and the fiscal obligations of that state. The plan is designed to concentrate NGEU spending in lower-income economies and those that suffered most from the pandemic shock.

The Next Generation EU program builds on the long-standing EU common budget. However, the common budget has been stable at around 1 percent of Europe's GDP for many years on essentially a balanced budget basis, with no countercyclical component. In contrast, the NGEU will support a temporary but significant increase in spending deployed over 2020–2026. Moreover, the NGEU will be financed by EU-level debt issuance that will be repaid over a long horizon up to 2058. Taken together, the SURE and NGEU programs envisage new debt issuance by the European Union to the tune of €850 billion (about 6.0 percent of EU GDP): this corresponds to a sixteen-fold increase in the stock of EU debt, which stood at only €53 billion at the end of 2019. In short, the NGEU constitutes an EU-level macroeconomic stabilization instrument, by enabling a debt-financed increase in the EU budget in response to the pandemic shock.

The European Union does not have a central tax collection agency: all taxes are collected by the member states. Since EU-level debt is commonly backed by the member states, the grant component of the Next Generation EU program will not raise the national debt level of the member states, while its loan component will enable those member states with higher borrowing costs to fund national deficits at a cheaper rate. Just as the standard EU budget is financed by a mix of EU-dedicated funding streams (equivalent to federal taxes) and national contributions, the servicing of the NGEU debt will involve a mix of new EU-level dedicated revenue streams and extra national contributions. The expansion of EU-level dedicated taxes is another fundamental characteristic of a deeper EU-level fiscal union.

The Next Generation EU instrument was designed as a temporary initiative in response to the pandemic shock. However, it suggests that joint EU-level fiscal initiatives can be envisaged for large future common shocks. Through these various initiatives—the European Stability Mechanism, the SURE employment preservation loans, and the Next Generation EU initiatives—the stock of commonly-backed euro bonds could increase by about 9.6 percent of euro-area GDP. One can imagine a hybrid response emerging in the future: national fiscal policies de facto would be assigned the lead countercyclical role in relation to asymmetric and/or minor shocks, with a scaled-up EU-level fiscal capacity standing ready to address tail risks. The alternative interpretation is that the pandemic is a truly exceptional event, with the vast bulk of debt issuance likely to remain at the national level.

A large stock of jointly backed debt would improve the resilience of the European monetary union through several mechanisms. In particular, it would reduce the likelihood of national financial instability within the euro area, by which an adverse shock is amplified by an increase in the sovereign risk premium. This

pro-cyclicality can be especially acute in a monetary union, because investors can easily switch to other sovereign bonds without taking on currency risk. An area-wide safe asset would facilitate a more integrated financial system by providing a common benchmark for asset pricing and liquidity management by banks; it would also make the euro more popular as an international reserve currency.

A permanent increase in the size of the EU-level budget (going beyond the temporary Next Generation EU initiative) is not the only route to expand the stock of area-wide safe assets (for an overview, see Leandro and Zettelmeyer 2018). There are also a range of proposals by which the stock of commonly backed debt would be expanded even if spending and taxation decisions remain at the national level, with the commonly backed debt used to make loans to the member countries. To ensure that the commonly backed debt is highly rated, these schemes typically require that national sovereign bonds would de facto be treated as subordinated to the common bonds. An alternative option in expanding the stock of area-wide safe assets would be to assemble pooled portfolios of national sovereign bonds in order to issue tranches of area-wide bonds, with the lowest-risk senior tranche constituting European Safe Bonds (ESBies) (for discussion, see Brunnermeier et al. 2016; Brunnermeier et al. 2017; High-Level Task Force on Safe Assets 2018). While securities backed by sovereign bonds would represent a significant financial innovation, it would be an effective device to improve the operation of a financial system that combines a single currency with 19 sovereign bond markets, especially if the prospects for greater commonly backed debt issuance are limited (Alogoskoufis and Langfield 2020).

The openness of the EU member countries to an expansion of area-wide borrowing remains an open question: after all, it would alter the status of their national sovereign bonds in financial markets. The challenge here may be to design a system that supports both more stable public finances at the national level while also embracing the complementary value of a joint liability as a shield against future asymmetric shocks.

Looking ahead, it is an open question whether the Next Generation EU initiative constitutes a case study of how a systemic shock can trigger a leap forward in fiscal integration. The full reform agenda for the euro area has been analyzed in a range of studies, including Juncker et al. (2015), Corsetti et al. (2015), Corsetti et al. (2016), and Bénassy-Quéré et al. (2018). The future pace of reform and the ultimate steady state for the euro area will turn on the evolving political economy of the euro area—which is the topic of the next section.

The Euro as a Political Project

The political foundations of the euro are extensively documented: the political leaders of the time intended to use the single currency as a mechanism to reinforce the integration process among the EU member countries, especially in the wake of German re-unification (Sandbu 2015; Brunnermeier, James, and Landau 2016; Mody 2018). As discussed by Eichengreen (1996), the political economy case for

monetary union could also be motivated as a mechanism to strengthen the political sustainability of deep economic integration in the European Union. However, the corresponding risk that a single currency would raise political tensions among the member states, rather than reducing them, was clearly identified in the 1990s debate on the desirability of a European monetary union (in this journal, Feldstein 1997). Indeed, Guiso et al. (2019) find that the constraints of the single currency have added to the rise of populism in some member countries.

Since banking systems and fiscal policies remained at the national level in the original institutional framework for the euro, the resilience of the euro depended on the willingness of the member states to commit to deeper area-wide integration if (or when) it turned out that the original design proved insufficiently comprehensive to protect macro-financial stability (James 2012; Spolaore 2013; Guiso, Sapienza, and Zingales 2016). Accordingly, from the beginning, the resilience of the euro has been ultimately underpinned by the intertwined nature of the support among the member states for the monetary union and the wider EU institutional framework.

The codependence between the resilience of the euro and the resilience of the European Union may not be fully visible to a casual observer, even if the euro is defined as the "currency of the European Union" in the Treaty of Europe. A basic reminder of the interlinked prospects for the euro and the European Union is the steady entry of new EU member states into the euro: it is planned that the euro area will expand in the coming years to 21 members (with the entry of Bulgaria and Croatia), up from the original eleven members. Accordingly, the political calculus about providing institutional backing for the euro reflects the strong joint political commitment to the European Union.

By design, the European Union is far less integrated than a federal system of national government like the United States. At the same time, the extent of shared sovereignty in many policy spheres among the EU member states is far more extensive than any other regional political grouping. In terms of economic integration, the EU single market is underpinned by the "four freedoms" governing the unrestricted movement of goods, persons, services, and capital (discussed further in this symposium by Head and Mayer). The single currency constitutes the most advanced type of shared sovereignty among the member countries—the analytical headaches arise in assessing how a single currency operates when other policy instruments remain primarily in the control of the member states.

This hybrid arrangement is inherently fuzzy. It represents a unique approach to address the various policy trilemmas that characterize international economic activity. For example, the "international monetary policy trilemma" refers to the impossibility of having all three these policies: fixed exchange rates, international capital mobility, and independent monetary policies (Obstfeld, Shambaugh, and Taylor 2005).[7] The "financial trilemma" refers to the impossibility of combining

---

[7]:In the European context, the international monetary trilemma is sometimes extended to refer to an "inconsistent policy quartet" whereby the political sustainability of high levels of free trade is also called into doubt by exchange rate volatility (Padoa-Schioppa 2004).

the three policies of integrated financial systems, international capital mobility, and independent financial stability policies (Schoenmaker 2011). The "international political economy trilemma" refers to the incompatibility of international economic and financial integration, transnational governance, and autonomous national democratic systems (as discussed in this journal by Rodrik 2000). The euro, the "four freedoms," and the overall EU project can be viewed as trying to find a middle path between full transnational governance and restricting the boundaries of policy regimes just to nation states (as discussed in Lane 2019a).

Compared to other regions, the European historical context, geographic proximity, and the scope for a high degree of economic and financial integration means that a hybrid governance system has evolved. It is composed of a mix of shared sovereignty in some areas and national autonomy in other areas, which does not fit neatly into either the category of a federal system or a loose alliance of nation states. The continuous testing of the perimeters of shared sovereignty defines the cut and thrust of European political debate.

Moreover, the codependence between the euro and the European Union is more binding now than before the euro was created, because an unravelling of the euro would almost surely have an adverse impact on the wider scope for political cooperation across the full spectrum of EU activities. The interconnectedness of the euro and the European Union has been further reinforced by Brexit. The United Kingdom was a member of the European Union but not of the euro zone, and thus Brexit means that the overlap between EU and euro area membership is now greater than ever (and will further expand with the planned entry of Bulgaria and Croatia to the euro area in the near future).

Even more directly, the strengthening of the euro's institutional framework since 2012 means that the economic and financial systems of the member states are far more intertwined today than in the first decade of the euro (including a greater role for shared public backstops), which strengthens the common interest in the ongoing success of the single currency. The level of fiscal integration has been further expanded in response to the pandemic emergency, through a scaling up of the issuance of EU-level debt and a greater role for EU-level spending programs.

The current pandemic shock is an important test case for the euro, given the greater scope for a common monetary policy to manage effectively a common shock, compared to solely national-level policies. The pandemic can be interpreted as a blend of a common shock and an asymmetric shock (because it has affected the EU member states in a non-uniform way). The common component of the pandemic shock triggered a significant monetary policy response from the European Central Bank and innovative fiscal measures at the EU level. The capacity of the euro area to absorb purely asymmetric shocks will surely be tested again in the future.

Conclusion

Perspectives on the euro inevitably require periodic updating: the capacity of the monetary union to withstand the severe challenges posed by the 2008–2013 crises (including through considerable reforms of its institutional architecture) and the relative stability of the euro area in recent years suggests that the euro has proven to be more resilient than was feared by many analysts during the darkest phases of the crisis period. The improvement in the euro area economy has been accompanied by a rise in the popularity of the euro in opinion surveys: 76 percent of respondents supported the euro in November 2019, which is a substantial improvement compared to the 67 percent support in the trough of the crisis in 2012 (see European Commission Eurobarometer and Bergbauer et al. 2020).

At the same time, the pandemic crisis poses new challenges for the euro area. The initial monetary and fiscal responses across the euro zone have been vigorous, and there has been a step increase in the extent of fiscal integration. However, building the consensus required to greenlight the further reforms that would enhance the resilience of the euro remains an ongoing challenge.

References

Alogoskoufis, Spyros, and Sam Langfield. 2020. "Regulating the Doom Loop." *International Journal of Central Banking* 16 (4): 251–92.

Arnold, Nathaniel G., Bergljot Barkbu, H. Elif Ture, Hou Wang, and Jiaxiong Yao. 2018. *A Central Fiscal Stabilization Capacity for the Euro Area.* IMF Publications.

Bastasin, Carlo. 2012. *Saving Europe: Anatomy of a Dream.* Washington, DC: Brookings Institution Press.

Bénassy-Quéré, Agnès, Markus K. Brunnermeier, Hendrik Enderlein, Emmanuel Farhi, Marcel Fratzscher, Clemens Fuest, Pierre-Olivier Gourinchas et al. 2018. *Reconciling Risk Sharing with Market Discipline: A Constructive Approach to Euro Area Reform.* Centre for Economic Policy Research.

Bergbauer, Stephanie, Nils Hernborg, Jean-Francois Jamet, Eric Persson and Hanni Schölermann. 2020. "Citizens' Attitudes towards the ECB, the Euro and Economic and Monetary Union." ECB Economic Bulletin 4/2020.

Berger, Helge, Giovanni Dell'Ariccia, and Maurice Obstfeld. 2018. "Revisiting the Economic Case for Fiscal Union in the Euro Area." IMF Departmental Paper 18/03.

Bianchi, Javier, and Jorge Mondragon. 2019. "Monetary Independence and Rollover Crises." Unpublished.

Brunnermeier, Markus K., Luis Garicano, Philip R. Lane, Marco Pagano, Ricardo Reis, Tano Santos, David Thesmar, Stijn Van Nieuweburgh and Dimitri Vayanos. 2016. "The Sovereign-Bank Diabolic Loop and ESBies." *American Economic Review* 106 (5): 508–12.

Brunnermeier, Markus K., Sam Langfield, Marco Pagano, Ricardo Reis, Stijn Van Nieuweburgh, and Dimitri Vayanos. 2017. "ESBies: Safety in the Tranches," *Economic Policy* 32 (90): 175–219.

Brunnermeier, Markus K., Harold James, and Jean-Pierre Landau. 2016. *The Euro and the Battle of Ideas.* Princeton, NJ: Princeton University Press.

Brunnermeier, Markus K., and Ricardo Reis. 2019. "A Crash Course on the Euro Crisis." NBER Working Paper 26229.

Carmassi, Jacopo, Sonja Dobkowitz, Johanne Evrard, Laura Parisi, André Silva, and Michael Wedow. 2020. "Completing the Banking Union with a European Deposit Insurance Scheme: Who is Afraid of Cross-Subsidisation?" *Economic Policy* 35 (101): 41–95.

Corsetti, Giancarlo, Keith Kuester, André Meier, and Gernot J. Müller. 2014. "Sovereign Risk and Belief-Driven Fluctuations in the Euro Area." *Journal of Monetary Economics* 61: 53–73.

Corsetti, Giancarlo, Lars P. Feld, Philip R. Lane, Lucrezia Reichlin, Hélène Rey, Dimitri Vayanos, and Beatrice Weder di Mauro. 2015. *A New Start for the Eurozone: Dealing with Debt.* Centre for Economic Policy Research.

Corsetti, Giancarlo, LarsP. Feld, Ralph Koijen, Lucrezia Reichlin, Ricardo Reis, Hélène Rey, and Beatrice Weder di Mauro. 2016. *Reinforcing the Eurozone and Protecting an Open Society.* London: Centre for Economic Policy Research.

Corsetti, Giancarlo, Barry Eichengreen, Galina Hale and Eric Tallman. 2020. "The Euro Crisis in the Mirror of the EMS: How Tying Odysseus to the Mast Avoided the Sirens but Led Him to Charybdis." *Open Economies Review* 3: 219–36.

Draghi, Mario. 2012. "Speech by Mario Draghi, President of the European Central Bank at the Global Investment Conference in London 26 July 2012." https://www.ecb.europa.eu/press/key/date/2012/html/sp120726.en.html.

Eichengreen, Barry. 1996. "A More Perfect Union? The Logic of Economic Integration." Princeton Essays in International Finance 198.

Eichengreen, Barry. 2010. "The Breakup of the Euro Area." In *Europe and the Euro*, edited by Alberto Alesina and Francesco Giavazzi, 11–51. Chicago: University of Chicago Press.

Farhi, Emmanuel, and Iván Werning. 2016. "A Theory of Macroprudential Policies in the Presence of Nominal Rigidities." *Econometrica* 84 (5): 1645–1704.

Feldstein, Martin. 1997. "The Political Economy of the European Economic and Monetary Union: Political Sources of an Economic Liability." *Journal of Economic Perspectives* 11 (4): 23–42.

Fernández-Villaverde, Jesús, Luis Garicano, and Tano Santos. 2013. "Political Credit Cycles: The Case of the Eurozone." *Journal of Economic Perspectives* 27 (3): 145–166.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2016. "Monnet's Error?" *Economic Policy* 31 (86): 247–97.

Guiso, Luigi, Helios Herrera, Massimo Morelli, and Tommaso Sonno. 2019. "Global Crises and Populism: The Role of Eurozone Institutions." *Economic Policy* 34 (97): 95–139.

Hartmann, Philipp, and Frank Smets. 2018. "The European Central Bank's Monetary Policy During Its First 20 Years." *Brookings Papers on Economic Activity* Fall (2018): 1–146.

High-Level Task Force on Safe Assets. 2018. *Sovereign Bond-Backed Securities: A Feasibility Study.* European Systemic Risk Board.

Honohan, Patrick. 2019. *Currency, Credit and Crisis: Central Banking in Ireland and Europe.* Cambridge, UK: Cambridge University Press.

James, Harold. 2012. *Making the European Monetary Union.* Cambridge, MA: Harvard University Press.

Juncker, Jean-Claude, Donald Tusk, Jeroen Dijsselbloem, Mario Draghi, and Martin Schulz. 2015. *Completing Europe's Economic and Monetary Union.* Brussels: European Commission.

Lane, Philip R. 2006. "The Real Effects of European Monetary Union." *Journal of Economic Perspectives* 20: 47–66.

Lane, Philip R. 2012. "The European Sovereign Debt Crisis." *Journal of Economic Perspectives* 26 (3): 49–68.

Lane, Philip R. 2019a. "Globalisation: A Macro-Financial Perspective." *Economic and Social Review* 50 (2): 249–63.

Lane, Philip R. 2019b. "Macrofinancial Stability and the Euro." *IMF Economic Review* 67 (3): 424–42.

Lane, Philip R. 2019c. "Determinants of the Real Interest Rate." Speech, National Treasury Management

Agency, Dublin, November 28, 2019. htttps://www.ecb.europa.eu/press/key/data/2019/html/ecb.sp191128_1>de8e7283e6.en.html.

Leandro, Alvaro, and Jeromin Zettelmeyer. 2018. "The Search for a Euro Area Safe Asset." PIIE Working Paper 18-3.

Martin, Philippe, and Thomas Philippon. 2017. "Inspecting the Mechanism: Leverage and the Great Recession in the Eurozone." *American Economic Review* 107 (7): 1904–37.

Marzinotto, Benedicta, André Sapir, and Guntram B. Wolff. 2011. "What Kind of Fiscal Union?" Bruegel Policy Brief 2011/06.

Mody, Ashoka. 2018. *Euro Tragedy: A Drama in Nine Acts.* Oxford: Oxford University Press.

Mundell, Robert A. 1961. "A Theory of Optimum Currency Areas." *American Economic Review* 51 (4): 657–65.

Obstfeld, Maurice, Jay C. Shambaugh, and Alan M. Taylor. 2005. "The Trilemma in History: Tradeoffs among Exchange Rates, Monetary Policies, and Capital Mobility." *Review of Economics and Statistics* 87 (3): 423–38.

O'Rourke, Kevin H., and Alan M. Taylor. 2013. "Cross of Euros." *Journal of Economic Perspectives* 27 (3): 167–92.

Padoa-Schioppa, Tommaso. 2004. *The Euro and Its Central Bank: Getting United After the Union.* Cambridge, MA: MIT Press.

Rodrik, Dani. 2000. "How Far Will International Economic Integration Go?" *Journal of Economic Perspectives* 14 (1): 177–86.

Rostagno, Massimo, Carlo Altavilla, Giacomo Carboni, Wolfgang Lemke, Roberto Motto, Arthur Saint Guilhem, and Jonathan Yiangou. 2019. "A Tale of Two Decades: The ECB's Monetary Policy at 20." ECB Working Paper 2346.

Sandbu, Martin. 2015. *Europe's Orphan: The Future of the Euro and the Politics of Debt.* Princeton, NJ: Princeton University Press.

Schelkle, Waltraud. 2017. *The Political Economy of Monetary Solidarity: Understanding the Euro Experiment.* Oxford: Oxford University Press.

Schoenmaker, Dirk. 2011. "The Financial Trilemma." *Economics Letters* 111 (1): 57–59.

Spolaore, Enrico. 2013. "What is European Integration Really About? A Political Guide for Economists." *Journal of Economic Perspectives* 27 (3): 125–44.

Whelan, Karl. 2019. "The Euro at 20: Successes, Problems, Progress and Threats." *Economic and Social Review* 50 (4): 725–50.

The United States of Europe: A Gravity Model Evaluation of the Four Freedoms

Keith Head and Thierry Mayer

The idea of creating a "United States of Europe" has a long history of hopes outstripping achievements. The first well-known statement was by Victor Hugo (1849). Speaking at an International Peace Congress, the author-turned-politician prophesied: "A day will come when these two immense groups, the United States of America and the United States of Europe, shall be seen placed in presence of each other, extending the hand of fellowship across the ocean." He also forecast that inter-European wars would end: "A day will come when the only fields of battle will be markets opening up to trade and minds opening up to ideas." In the century after Hugo's speech, Germany and France went to war three times. At the end of the third of these wars, Winston Churchill (1946) repeated the call to "build a kind of United States of Europe."

The concept of a United States of Europe encompasses three different policy objectives. First, it expresses the wish to end the wars that plagued the continent for centuries. Second, it embodies the hope to unify a market as large and deeply integrated as its counterpart across the Atlantic. Finally, there is the goal of a political union: a subordination of the original nation states under a federal government. What progress has been made towards each of these objectives?

■ *Keith Head is HSBC Professor in Asian Commerce, Sauder School of Business, University of British Columbia, Vancouver, Canada. He was María de Maeztu Visiting Professor at CEMFI (Center for Monetary and Financial Studies), Madrid, Spain, during the 2019–2020 academic year. He is also a Research Fellow at the Centre for Economic Policy Research in London, United Kingdom. Thierry Mayer is Professor of Economics, Sciences Po, Paris, France. He is also a Scientific Adviser, Centre d'Études Prospectives et d'Informations Internationales (CEPII), Paris, France, and a Research Fellow, Centre for Economic Policy Research, London, United Kingdom. Their emails are keith.head@sauder.ubc.ca. and thierry.mayer@sciencespo.fr.*

The current European Union has its origins in the European Coal and Steel Community. The French foreign minister Robert Schuman (1950) explicitly prioritized the peace objective declaring, "The solidarity in production thus established will make it plain that any war between France and Germany becomes not merely unthinkable, but materially impossible." While the experience of the US Civil War cautions against complacency, the European unification project appears to have succeeded at ending war between its members. The Palmer et al. (2015) military dispute data show that there has never been use of force between contemporaneous members of the European Community.

Progress on the third policy objective, political union, has been assessed in two articles in this journal. Feldstein (1997) construes the move towards monetary union as "a way to further the political agenda of a federalist European political union, which will have a common foreign and military policy and a much more centralized . . . economic and social policies." Among other negative consequences, Feldstein predicted that declining competition within the European Union would lead to rising protectionist measures against non-European countries. Seven years later, Alesina and Perotti (2004) open their paper by flatly stating that Europe is not "building a federal state similar to the United States." They argue that deficient institutions and incompatible goals constrain the path of European unification. With these and other disadvantages stacked against the European integration project, there would appear to be little chance of fulfilling the greatest aspirations of Euro-optimists. Of course, Euro-optimism was not universal: the sovereignty concerns voiced by Brexit advocates show that Feldstein was far from alone in viewing greater political centralization as anathema.

The first objective, peace, seems settled; the third, political union, seems remote. What of the second objective, expressed in Hugo's vision of "markets opening to trade"? This goal of reciprocal market openness is an alternative way to envision a United States of Europe. The 1957 Treaty of Rome set out the commitment of the member states to the four freedoms of movement: goods, services, persons, and capital. In this essay, we ask if Europe is approaching the levels of economic integration in terms of these four freedoms found amongst the 50 states of the United States. We report here—with some degree of surprise—a body of quantitative evidence suggesting that, by several important metrics, European states have matched or surpassed the levels of openness prevailing amongst the 50 American states. Furthermore, we find that increased integration within Europe has come from lower intra-European barriers, rather than the rise of a "Fortress Europe" excluding external flows.

We begin with a primer on the phases of European economic integration. Remembering the timing will be useful in interpreting the evidence from a gravity model on flows of goods, people, and capital amongst European countries, as well as some evidence on convergence of prices and incomes.

A Primer on European Economic Integration Policies

The policies fostering EU economic integration can be usefully divided into three phases. The first phase is the original decade-long implementation of the

1957 Treaty of Rome that called for free movement of goods, services, persons, and capital. The policies of the second phase, the Single Market Program (SMP) spanning from 1986 to 1992, were intended to reduce the remaining border-related, non-tariff barriers to all four movements. The third phase began when the Maastricht Treaty entered into force at the end of 1993. The Maastricht Treaty created the European Union along with the European Central Bank and the beginning of movement toward the euro as a single currency. It also created a number of non-economic European institutions, including a common foreign and security policy and cooperation in justice and security.

For the free movement of goods, Article 3a of the Treaty of Rome created a timetable for elimination of tariffs and quantitative restrictions by 1970. Prior to the Treaty of Rome, the six signatories had sizeable tariffs: about 16 to 20 percent for France, Germany, and Italy, and 10 percent for Belgium, the Netherlands, and Luxembourg (1955 unweighted averages from Bown and Irwin 2016, Table 3). Tariffs were eliminated in 1968, which is also the year when the European Community became a customs union with a common external tariff.

By the early 1980s, it had become evident that the removal of formal trade impediments had not fully integrated the European market. Physical borders still impeded the flow of goods, and national product regulations still had the effect of shutting out imports. Prominent examples of the latter included Italy's requirement that all pasta contain 100 percent durum semolina and Germany's law mandating that beers include just the four ingredients authorized by the 1815 Bavarian Purity Law. While both policies would be struck down by the European Court of Justice, a White Paper by the European Commission (1985) titled "Completing the Internal Market" pointed to a broader need to remove technical barriers to trade. It listed 300 measures to implement deeper integration via elimination of frontiers, mutual recognition, and harmonization of regulations. These measures, legislated by the 1986 Single European Act would be implemented by the end of 1992.

Because the provision of services across national borders is often realized through foreign affiliates or embedded in professionals who travel in order to deliver the service, it is not surprising that the three freedoms other than goods—persons, services, and capital—were grouped by the writers of the original Treaty of Rome in Article 3c. Free movement of factors was not just instrumental for liberalizing service trade, it also became a goal in its own right. While the Single Market Program included some services measures, further liberalization would not come until the 2006 "Services Directive" to facilitate cross-border trade. Progress remained slow, leading Brussels to bring a lawsuit against all 28 members for failure to comply (as reported by Brunsden 2019).

Regarding migration, the Treaty of Rome's Article 48 committed members to ensure "free movement of workers . . . within the community" by the end of a 12-year transition period (which would have been in 1970). This provision gave workers the right to travel within Europe in search of employment and it prohibited discrimination on the basis of nationality. Further treaties tried to address some of the remaining impediments to migration within the Union. The 2007 Treaty of Lisbon extended free movement of workers to cover all persons, like retirees.

The Treaty of Rome also called for member states to abolish restrictions on capital flows amongst themselves. Baldwin and Wyplosz (2019) report that little progress was made in this area over the following 30 years, mainly because of loopholes within the Treaty. Article 73 allowed for "protective measures" on capital flows when needed to "avoid disturbances in the functioning of the capital market." When such measures were deemed necessary, member states could act without prior authorization from the European Commission. For example, France asserted the right to deny approval for foreign direct investments above certain thresholds (set at 50 million French francs). The Single Market Program and the Maastricht Treaty finally brought about the promised liberalization in the early 1990s. Starting in 1996, France had to allow all inward investment from EU members and could only apply the restrictions to investors outside the Union (for details, see https://www.senat.fr/rap/l95-191/l95-191_mono.html).

A Gravity Approach to Measuring Economic Integration

Using a gravity model to estimate the impact of trade agreements on bilateral flows goes back to Tinbergen (1962). The moniker comes from an analogy to Newton's gravity where the mass of two objects is replaced with the size of two economies, and trade volumes replacing force exerted. The attraction of the gravity model for our purposes goes beyond historical practice. Modern gravity models allow us to estimate the underlying costs of cross-border movement for all four freedoms in a unified framework.

The crucial insight is that each type of flow can be thought of as the outcome of a discrete choice problem. For goods and services, the choice is between source countries for a given product (as in Eaton and Kortum 2002 and Head, Mayer, and Ries 2009, respectively). For migrants, the choice is the country in which to reside (for an early logit model, see Tabuchi and Thisse 2002; for an application, see Grogger and Hanson 2011). Finally, for capital, we imagine an asset owner seeking to sell to the highest bidder (as in Head and Ries 2008). All these choices can be formalized as an agent selecting the option with the maximum $(A_i/\tau_{ni})z_{hi}$, where A_i captures the attractiveness of option i to all agents, τ_{ni} captures the costs of separation between agents located in n and i, and z_{hi} captures the idiosyncratic factors influencing agent h's choice. With an assumption about the distribution of z (specifically, it comes from a Frechet distribution, with dispersion parameter ϵ), the probability for all agents in n choosing i has the following form:

$$\pi_{ni} = \frac{(A_i/\tau_{ni})^\epsilon}{\sum_j (A_j/\tau_{nj})^\epsilon}.$$

Getting to a modern gravity model that can be estimated with data takes four more steps. First, the total flow (of goods, migrants, and so on) can be obtained by multiplying the fraction π_{ni}, by the mass of country n's choosers, X_n; that is, $X_{ni} = \pi_{ni} X_n$. The second step is to utilize fixed effects to absorb the country-specific terms: A_i, X_n, and the n-specific denominator in the equation above. Third,

computation becomes more transparent once we realize this structure is (like Newton's gravity equation) linear in logs. That is, $\ln X_{ni}$ is linear in i and n fixed effects and $\ln \tau_{ni}$. This last term has an elasticity $-\epsilon$ that Arkolakis, Costinot, and Rodríguez-Clare (2012) established to govern the impact of trade costs on economic welfare. The final step, when we move to panel data, is to allow for the characteristics of each country (A_{it}) and the frictions (τ_{nit}) to vary over time.

For goods and services we follow the convention of referring to τ_{nit} as "trade costs"; when speaking of all four movements, we refer to "frictions." The interpretation of τ_{nit} depends on the movement under consideration. For goods, the obvious factors creating trade costs are tariffs and transport costs. For services, there are regulatory restrictions and travel costs for in-person services. For migration, the relevant τ_{nit} determinants are the transferability of human capital (such as acceptance of educational credentials) and the cost of maintaining connections to the origin country. For foreign asset-ownership, there are regulatory impediments as well as the costs of remote management emphasized in Head and Ries (2008). Considering each of these cases, some frictions are continuous functions of the distance, broadly defined, between i and n, but others rise discontinuously at the border. We will control for distance and measure integration of the overall market by its impact on border frictions.

We use a long panel approach to evaluate the success of Europe's policies that have sought to bring about lower impediments to each of the four movements. For trade and migration we can take the data back to 1960, just two years after the Treaty of Rome was implemented. Estimating changes in τ using this long panel has two attractive aspects. First, we can control for all unobserved linkages between country pairs that persist over time by using fixed effects. Second, we can compare the timing of the evolution of estimated τ_{nit} to the timing of single market policies recounted in the previous section.

Studies estimating the effect of trade agreements on τ_{nit} typically specify $-\epsilon \ln \tau_{nit} = \beta \, EU_{nit} + v_{nit}$, where $-\epsilon$ is again the key elasticity that governs the effect of changes in trade costs on welfare. In this specification, $EU_{nit} = 1$ if i and n are both members of the EU year t and v_{nit} comprises other determinants of frictions. However, in this paper three distinct concerns motivate us to develop a richer specification of EU effects. First, back to the primer in the previous section, EU integration progressed in phases and did not simply turn on a constant level in a certain year. Therefore we allow for time-varying effects of EU membership; that is, we estimate the β_t coefficient on the EU variable for every year. The second concern is that EU membership obviously does not come about through randomized assignment. It is difficult to imagine what quasi-random variation could be exploited to estimate causal effects of EU membership. Following Baier and Bergstrand (2007), we control for a first-order source of endogeneity: unobservable bilateral frictions and linkages (linguistic and cultural similarity, distance and aspects of physical geography like the English Channel or the Alps). We specify that v_{nit} includes dyad-ni fixed effects to control for such time-invariant factors, so as to identify EU effects from the long-run evolution of bilateral flows following each members' accession to the European Union.

The first two solutions, time-varying β_t and dyadic fixed effects, were employed in Mayer, Vicard, and Zignago (2019) to study EU effects on trade. Here we address a third concern—applicable to all four freedoms—that becomes apparent only after examining carefully the structure underlying the gravity model. The earlier equation showing the probability of agents in country n choosing i shows us that only relative trade costs matter. If we scale up all the τ, people will be poorer but the movement decisions will not change. There could be strong flows between two countries, say France and Germany, because they impose low barriers on each other, or because they impose high barriers on everyone else (the "fortress Europe" scenario). The appropriate measure of whether EU integration is working is not whether it causes the French to buy more German goods at the expense of fewer purchases of American goods, but whether the French buy more German goods in place of French goods. Estimating EU effects that are directly relevant for welfare therefore requires us to compare international flows to flows with self.[1]

To implement this solution, we distinguish three different types of flows as follows:

$$-\epsilon \ln \tau_{nit} = \beta_t^{EUB} \underbrace{B_{ni} EU_{nit}}_{\text{EU to EU}} + \beta_t^{CET} \underbrace{B_{ni}(1 - EU_{it})EU_{nt}}_{\text{ROW to EU}} + \beta_t^{ROW} \underbrace{B_{ni}(1 - EU_{nt})}_{\text{ROW imports}} + \upsilon_{nit},$$

where B_{ni} denotes the presence of a national border ($B_{ni} = 1 \Leftrightarrow n \neq i$). In this specification, $B_{ni}EU_{nit}$ indicates within-EU flows that cross a national border. The product $B_{ni}(1 - EU_{it})EU_{nt}$ captures the EU members' flows from third countries, where CET is a mnemonic for the "common external tariff" pertaining to trade flows into the region. Finally, $B_{ni}(1 - EU_{nt})$ corresponds to the flows to the rest of the world, denoted ROW. The standard EU trade effect—the net gains to EU consumers achieved by buying from an EU source instead of an outside country—corresponds to $\beta_t = \beta_t^{EUB} - \beta_t^{CET}$.[2]

The country-pair fixed effects in the panel gravity equations imply that our estimates can inform us about changes in τ_{nit}, but not give its level. In particular, dividing the EU border effect by the relevant friction elasticity and exponentiating (to undo the logarithmic transformation), we measure relative frictions as $\tau_{nit}/\tau_0 = \exp(-\beta_t^{EUB}/\epsilon)$. The baseline τ_0 is the first year of τ_{nit} for ROW destinations. The panel approach

---

[1] Here we make this comparison using regressions. An alternative approach uses the multiplicative structure of the equation in the text to *infer* what must be the impediments underlying an observed pattern of choices. These friction indices, derived by Head and Ries (2001) for trade flows, use flows to self in a way that makes it possible to distinguish the impediments to within-agreement flows from the ones imposed on nonmembers. As in Novy (2013), the inferred frictions can be regressed on determinants of trade costs. When frictions are symmetric, the Head-Ries index regressions and the gravity regression method (including self trade) yield the same results. However, preferential agreements lead to asymmetric frictions, which only the regression approach can handle appropriately.

[2] Online Appendix A available with this article at the *JEP* website, contains the full structure of the gravity model for goods and reveals the tariff and non-tariff barriers that underlie each of the β_t coefficients. A simulated version of the model demonstrates that the gravity regressions recover the barriers contained in the regression specification in the text. Figure A.1 in this Appendix illustrates the connection between the EU border effect estimate and the implied change in welfare from regional integration.

allows us to assess the progress the European Union has made relative to this benchmark, but it does not reveal the level of integration that the European Union has achieved. For that purpose, we will use in the next section a comparison of two cross-sections, the flow amongst US states and the flow amongst EU members. For goods, migration, and mergers, we can construct the interstate flow matrix in a way that is closely comparable to the international matrix within the European Union. We then can estimate the effect of crossing a border on each of these flows in each "union." Normalizing $\tau_{nn} = 1$, the border effect estimates the average $-\epsilon \ln \tau_{ni}$ for $n = i$. We can then divide the estimated β by an estimate of ϵ from the literature and exponentiate to obtain the implied ad valorem equivalent: AVE = $\exp(-\beta/\epsilon) - 1$. These cross-sectional estimates of the ad valorem equivalent of the border allow us to quantify the level of remaining impediments to movement in the European Union and in the next section compare directly to the United States, which we can think of as a plausible lower bound for impediments.

Details of the specific regressions and data we use are in the online Appendix available with this paper at the *JEP* website. All our regressions are estimated with the widely used Poisson pseudo-maximum likelihood estimation (PPML), including origin and destination fixed effects. When using panel data, the origin and destination effects vary over time, and we add dyadic fixed effects that absorb all time-invariant pair characteristics.[3] Here, we will use the share of total expenditure of the importing country as our dependent variable, which is consistent with using the level of trade (as proposed by Eaton, Kortum, and Sotelo 2013, and validated using Monte Carlo simulations by Head and Mayer 2014). Results using levels of trade are also provided in the online Appendix. For migration, the relevant denominator for the share variable is the population in the origin country. For asset transactions, we define shares as the ratios of n's acquisitions from country i relative to country n's total acquisitions. To implement our regression specification, we constructed a set of country-level flows with self for all four movements going back as far as 1960 for trade in goods and migration. While other datasets have created self-trade series—most recently the International Trade and Production Database for Estimation (ITPD-E) from the US International Trade Commission—those datasets have shorter time spans and do not include flows such as capital and migrants.

First Movement: Goods

Here, we evaluate the impact of six decades of efforts to facilitate trade in goods within the European Union. Figure 1 shows results obtained when estimating our augmented gravity equation on all the bilateral trade flows recorded from 1960 to 2018 (online Appendix B.2 describes how we constructed this data). The blue squares show the percentage change in implied trade costs (τ_{nit}) for trade between two members of the European Union relative to imports to the rest of world in 1960. In backing out the changes in τ from the left-hand side of the regression equation, we use the median of the literature's estimates for the critical elasticity

---

[3]Weidner and Zylkin (2019) prove consistency of the three-way fixed effect estimator under Poisson.

Figure 1
Estimates of the Evolution of Trade Costs in Europe: Goods
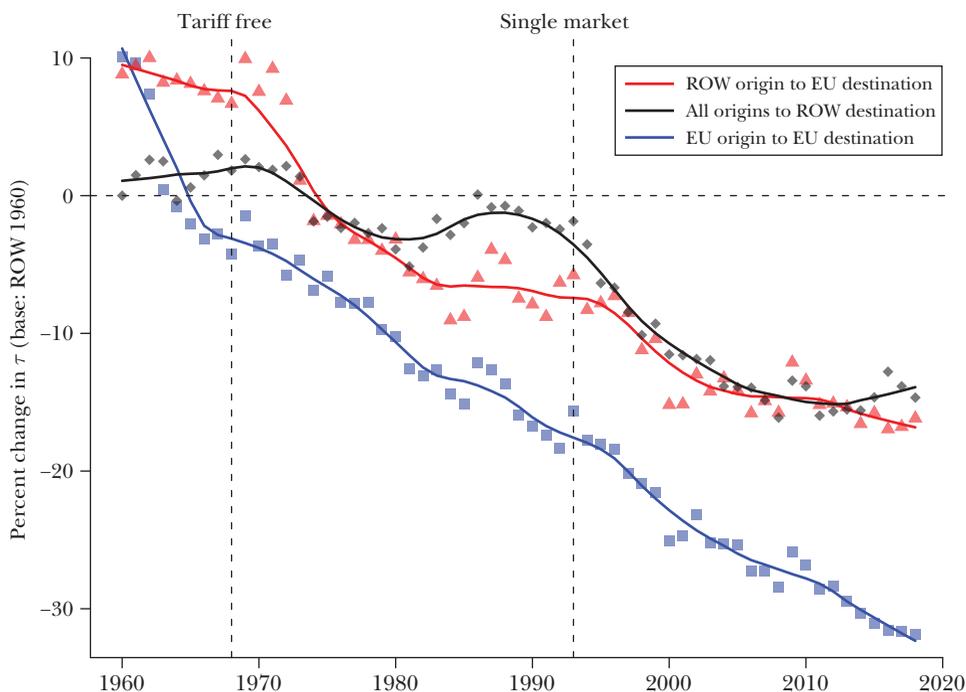
Source: See online Appendix B.2 for data sources.
Note: Each point is obtained by differencing with respect to the 1960 ROW-border coefficient, dividing by $-\epsilon = -5$, exponentiating, subtracting one, and multiplying by 100.

as collected by Head and Mayer (2014), $\epsilon = 5$. The red triangles show the trade costs EU members impose on their imports from the rest of the world. The black diamonds are the changes in trade costs of non-EU members (regardless of origin) relative to the base year. We can see that intra-EU trade started out 10 percent more costly than export to the rest of the world, then falls sharply during the period of tariff reductions, and then continues to decline until the present. The total decline is 38 percent. Trade costs by EU countries on outsiders have also fallen considerably, by 23 percent.

The striking finding here is that bringing self-trade into the estimation reveals a much larger trade-liberalizing trend than the traditional approach exemplified in Mayer, Vicard, and Zignago (2019). The worry was that the bias would go in the opposite direction: rising EU external protection could have been interpreted as rising internal integration. What the red triangles show is that the European Union actually lowered its external barriers.

Baier, Bergstrand, and Feng (2014) were the first to show that deep trade agreements take over a decade to realize their full integration gains, estimated to be about a doubling of trade. Limão (2016) summarizes the broader evidence on free trade agreements with a focus on deep integration agreements such as the European

Union. As we do here, he estimates the three-way fixed effect model. However, Limão estimates a single coefficient capturing the total effect of deep integration agreements. To compare our time-varying effects to his specification (which does not include internal trade), we take the difference in differences between the initial and the final coefficients for EU-EU and ROW-EU. In 1960 the two were nearly the same, but a large gap emerges by 2018, implying a standard EU effect estimate of 1.1, which amounts to a tripling of trade (that is, $\ln(3) = 1.1$). This is remarkably similar to the 1.2 coefficient estimated by Limão.[4]

The dashed vertical line in 1968 in Figure 1 shows the point at which tariffs ceased to be collected on intra-EU trade. In the years leading up to this, the steep fall in the blue squares shows the rapid progress towards integration inside the European Union. Even after tariffs had been eliminated, the persistent downward trend in the blue squares implies that internal liberalization continued at a steady pace until the present day. Starting in the early 1970s, the imports of EU member countries originating from non-EU countries also grew quickly. This attenuates the bias of EU members to trade within the bloc, which dampens the growth of the standard EU effect (the difference between the blue and the red estimates). The reason both types of trade can increase at once, even after controlling for origin-year fixed effects, is that EU members were trading less and less intra-nationally, which is the relevant criterion for welfare-improving regional liberalization.[5] The figure reveals declining intra-EU trade costs both before and after 1992 (shown with the second dashed line)—the year the single market program was completed. All in all, Figure 1 conveys an optimistic message of continuous progress for EU integration in goods.

Second Movement, Persons

While economists have been using gravity equations to estimate the trade effects of regional agreements since the 1960s, attention did not turn to the estimation of EU effects on migration until recently. However, development economists have long used gravity to study migration with a particular focus on migrant networks (for a survey, see Beine, Bertoli, and Moraga 2016). Important advances integrating trade and spatial economics (Redding 2016; Redding and Rossi-Hansberg 2017) show that the incentive to migrate depends in part on the openness to trade of the destination country. In a model featuring changes in both trade and migration costs, Caliendo et al. (2020) quantify the sources of welfare gains to the countries joining the European Union as part of the 2004 eastern enlargement. They report 30 percent of the gains come from trade policy, 68 percent from migration policy and 2 percent from the interaction of these two. For the European Union as a

---

[4]Using levels of trade instead of shares delivers a much lower difference in differences, as can be seen in online Appendix Figure C.2. The reason our Poisson in shares is more similar to Limão (2016) than our own Poisson in levels coefficient is that he uses a linear-in-logs specification. Eaton et al. (2013) show that Poisson in shares results are closer to the linear-in-logs specification. Head and Mayer (2014) explain this proximity in results using the underlying features of those estimators.
[5]Building on Costinot and Rodríguez-Clare (2014), section A.3.1 of the online Appendix explains why welfare effects of trade agreements move inversely with the share of intra-national trade.

whole, a calculation from Table 7 of Caliendo et al. (2020) reveals that the relative contributions of trade and migration are almost inverted, with 63 percent from trade and 35 percent from migration.

Here we estimate the EU effect on migration, following the same specification used for trade in goods. The literature on gravity equations in migration mainly uses migration flows as the dependent variable. The problem with migrant flows is that intra-national flow data are not widely available. Our regressions use United Nations and World Bank migration datasets on the number of residents by country of birth. "Migration to self" is the count of people who live in the country they were born in (calculated by subtracting the stock of immigrants from the total population).

Figure 2 depicts the changes in migration frictions implied by the gravity β_t coefficients. As with trade flows, we need an estimate for the key elasticity ϵ to do this. As there is currently no consensus value for the appropriate migration elasticity, we summarize the estimates from ten recent influential papers in online Appendix Table B.1. The starting point for this miniature meta-analysis of the migration elasticity is the review in this journal by Kleven et al. (2020) of estimates of the response of high earners to the highest income tax rates. We complement this line of research with a recent stream of work in economic geography. While the public finance literature seeks this elasticity to inform optimal taxation, the work in geography values the elasticity because of the critical role it plays in counterfactual policy experiments. The median estimate for a migration cost elasticity is $\epsilon = 1.63$.

As we saw with trade in goods, the implied migration impediments within the European Union started out more onerous than those imposed by the rest of the world. Over the 1960 to 2015 period, the implied frictions for intra-EU migration (depicted with blue squares) fall by 71 percent. Most of the reduction in intra-EU migration impediments occurs during the 1960s. This is in line with the rapid schedule of liberalized migration promised in the Treaty of Rome for its original members. Thereafter the pace of integration slackens. Intra-EU migration frictions in 2015 are no smaller than they were in 1990. Relative to 1960, the implied tax on immigration from non-members (the red triangle coefficients) falls by 37 percent. There is much less migration liberalization in the rest of the world. Non-EU migrant frictions (black diamonds) falls by just 16 percent. The results depicted here for migration reinforce the conclusion obtained with trade data: Fears of "Fortress Europe" were unfounded.

Third Movement, Services

Baldwin and Wyplosz (2019) note: "Even to this day, the tension between allowing EU members to take care of their own regulation of services has prevented truly free trade in services." Integration of the service sector has not only been harder to achieve in practice, but also harder to measure. The literature on services integration is correspondingly smaller, but two recent papers have estimated the EU effect. Using Eurostat data on tradeable services, Mayer, Vicard, and Zignago (2019) obtain a coefficient of 0.18. On a comparable sample of goods flows they estimate an

Figure 2
Estimates of the Evolution of Migration Frictions in Europe
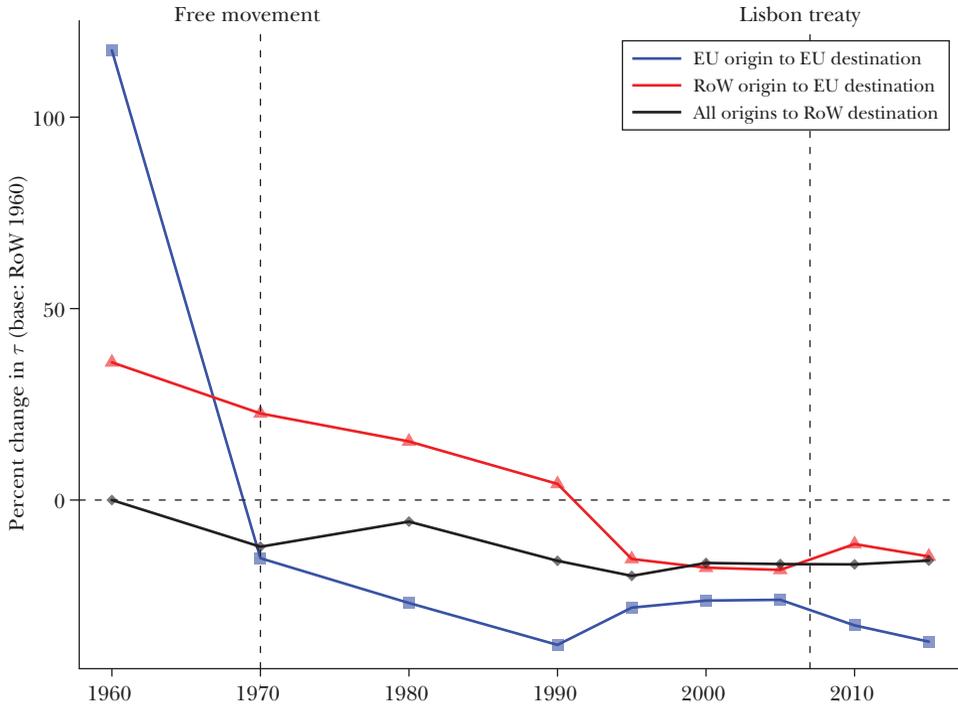
Source: See online Appendix B.3 for data sources.
Note: Each point is obtained by differencing with respect to the 1960 ROW-border coefficient, dividing by $-\epsilon = -1.63$, exponentiating, subtracting one, and multiplying by 100.

EU effect of 0.32. Heiland, Felbermayr, and Groeschl (2020) use the World Input-Output Database (with coverage from 2000 to 2014) and estimate larger EU effects for services (0.60) than for goods (0.43).

One of the distinctive advantages of the World Input-Output Database is that it contains intra-national trade flows, including for services. We therefore apply the same time-varying border effect analysis that we used for goods and migration to the case of services. This enables us to estimate the evolution of implied trade costs in services when a member of the EU imports from another member, when it imports from a third country, and when imports are from a rest-of-world country.[6]

---

[6] In contrast to Heiland, Felbermayr, and Groeschl (2020, table A2), we separate all the welfare-relevant border effects and let them evolve over time. Another distinction is that we keep only tradable services in the regression, dropping a number of International Standard Industry Classifications (construction, water distribution, electricity, sewerage, health, education and government services) that Heiland et al. attribute to services in the aim of running counterfactuals. With this change in industry coverage, the equivalent coefficients to the ones in table A2 of Heiland, Felbermayr, and Groeschl (2020) for goods and services are closer (0.43 and 0.48).

Figure 3 shows that frictions for services within the European Union are declining, as we have already seen for goods and migration. Again, to back out the changes in frictions, we need an estimate of the trade cost elasticity. Estimating this parameter for services is difficult because of the absence of tariffs and measurable transport costs. We therefore keep the same elasticity as we use for goods, which essentially assumes that the degree of heterogeneity in tastes or productivity revealed by this parameter is similar for goods and services. Using this $\epsilon = 5$, we can express the tariff equivalent of trade costs in services inside the European Union as being 11 percent smaller than when a non-EU country imports services in 2000. Over time, our estimates imply that those trade costs fell by 8 percent within the European Union. The European Commission, disappointed by the pace of reductions in barriers to service trade, issued a "Services Directive" in 2006. The data does not seem very impressed, and friction estimates continued on roughly the same trend line, perhaps explaining why the European Union recently brought a lawsuit for non-compliance against the 28 members. On a more positive note, we see from the red triangles that trade costs imposed on outsiders fell by 12 percent over the 2000–2014 period. Hence, we see no more evidence of "Fortress Europe" for services than for goods or migration.

Fourth Movement, Capital

Unlike the first three freedoms, where there is an obvious flow to use in gravity model estimation, there are multiple measures of capital flows we could consider, including foreign direct investment, portfolio investment, and mergers and acquisitions. Foreign direct investment is probably the most studied form of capital flow but there is no straightforward way to measure foreign direct investment to self (nor to measure foreign direct investment between states in the United States). Noting that Head and Ries (2008) find a 0.94 correlation between inward foreign direct investment and merger and acquisition flows for OECD countries, we see promise in examining mergers and acquisitions as a capital flow measure. The data easily allow for consistent calculation of merger and acquisition flows to self as well as the flows amongst American states that we will need later in the paper to compare levels of frictions.

Coeurdacier, De Santis, Aviat (2009) show that bilateral merger and acquisition flows are higher when both countries are members of the European Union. They use Securities Data Company (SDC) Platinum data to measure the flows of bilateral mergers and acquisitions but restrict attention to international transactions over the years 1985 to 2004. We also use SDC Platinum as the source for flows of mergers and acquisitions, but we extend it until 2018, and most importantly, we augment it with internal flows of mergers and acquisitions.

To convert the β_t coefficients into tax equivalents, we need the elasticity of capital flows with respect to taxes on cross-border movement. Ahern, Daminelli, and Fracassi (2015) estimate the elasticity of transaction values with respect to one plus the tax rate to be 5.03 (in column 1 of their Table 3). Coeurdacier, De Santis, and Aviat (2009) estimate a tax semi-elasticity of 4.46. Head and Mayer (2004) estimate a host-country tax semi-elasticity of 2.1 in a study of location choice of Japanese

Figure 3
Estimates of the Evolution of Trade Costs in Europe: Services
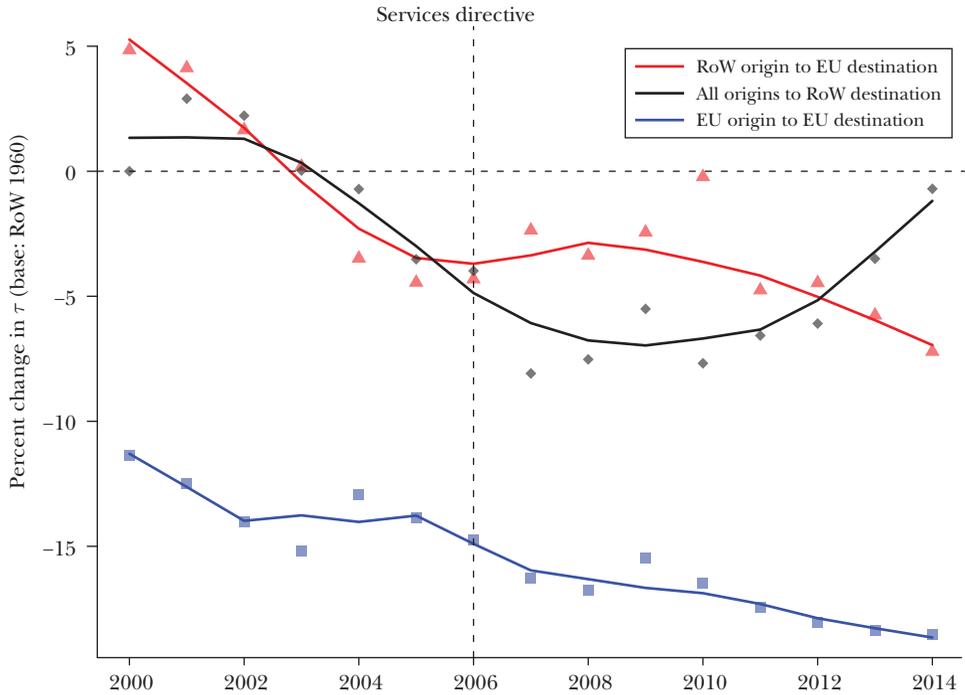
Source: See online Appendix B.4 for data sources.
Note: Each point is obtained by differencing with respect to the 1960 ROW-border coefficient, dividing by $-\epsilon = -5$, exponentiating, subtracting one, and multiplying by 100.

investments in Europe. We average these three estimates to obtain $\epsilon = 3.86$ as an estimate of the key elasticity.

Figure 4 depicts the time-varying border cost measured in ad valorem equivalents (AVEs) corresponding to EU acquirers of EU targets (blue squares), their non-EU acquirers of EU targets (red triangles), and transactions in the rest of the world (black diamonds). The normalization follows our practice in the other figures of using rest-of-the-world in the first year of the data. However, rest-of-the-world mergers and acquisitions in 1985 were very low compared to other years. The pattern we see in the figure is highly volatile from year to year so we smooth the data (using locally weighted smoothing). Over the 30 years of our data, EU members show little evidence of systematic bias in transactions against outsiders. However, our estimates imply EU targets are much less costly to acquire than targets in the rest of the world (the black diamonds). The absence of a clear upward trend in the blue squares from 1985 to 2018 tells us that membership in the European Union has not lowered the tendency of mergers and acquisitions activity to stay within borders of EU members, except possibly in the last decade.

Figure 4
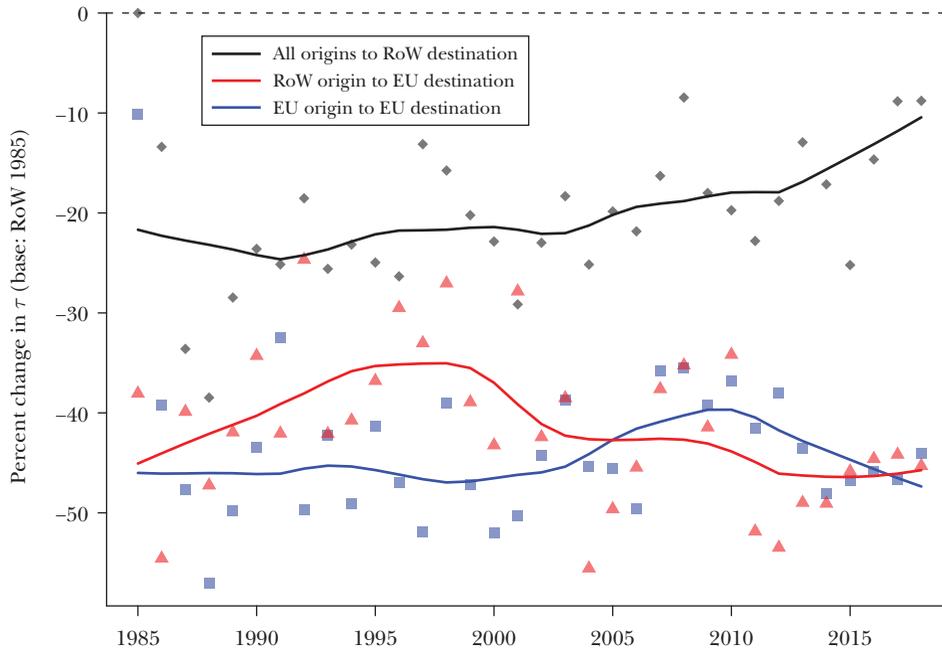**Estimates of the Evolution of M&A Frictions in Europe**

Source: See online Appendix B.4 for data sources.
Note: Each point is obtained by differencing with respect to the 1960 ROW-border coefficient, dividing by $-\epsilon = -3.86$, exponentiating, subtracting one, and multiplying by 100.

Comparing Levels of Border Barriers with the United States

We now compare flows inside the European Union and inside the United States to determine whether frictions within the 62-year old European Union have fallen to a level comparable to a federal state where many formal barriers have been banned for over two centuries. The regressions we discuss in this part of the paper are a simplified version of the ones from the previous section. Here, the estimation includes only intra-EU flows. Since $EU_{nit} = 1$ for all flows, and we estimate using cross-sectional data, our earlier regression equation collapses to $-\epsilon \ln \tau_{ni} = \beta B_{ni} + \upsilon_{ni}$. This leaves us with only one border coefficient to be estimated (β), which measures the tendency of EU countries to trade less with EU partners than with themselves. We then estimate an analogous equation where the flows are between and within the 50 American states. For both EU and US regressions, we control for distance (between and within the trading partners), but for the European Union we add a common language dummy. Head and Mayer (2010) show that border effects can be systematically overestimated if we use excessive distances to self. The Constant Elasticity of Substitution distance indexes used here should mitigate this problem. Even if some bias remains, we can still compare EU and US

border effects in the plausible case that approximately the same bias pertains in both areas.

For this investigation, we are limited to the three freedoms for which we can obtain comparable data for interstate flows in the United States (the EU part of the analysis retains the same datasets as in the previous section). Trade in goods within the 50 American states is measured using the Commodity Flow Survey (CFS). We use two issues of the CFS separated by 20 years, 1997 and 2017, and estimate a cross-section for each of those years. Migration data for the United States is constructed using the 2000 decennial census and the 2015 edition of the American Community Survey. Both provide a bilateral matrix of place of birth by state of current residence. As in the previous section, we use mergers and acquisitions transactions as our measure of capital movement, recognizing that it may not be representative of all types of capital. A very attractive feature of this flow is that SDC Platinum lists the state of both the acquirer and the target for almost all the transactions taking place in the United States. To the best of our knowledge, there is no US interstate data on trade in services.

Table 1 reports the tax equivalents of the border costs for each flow, region, and period. Recall that the ad valorem equivalent is calculated as $\text{AVE} = \exp(-\beta/\epsilon) - 1$, where β denotes the border coefficient. The gravity elasticities ϵ are the same ones from the previous section: 5.0 for trade, 1.6 for migration, 3.9 for mergers and acquisitions. For details of the regressions, including the raw coefficients on border, distance, and language, see online appendix Tables C.3, D.2, and F.1.

We begin with trade in goods. The data collection for the interstate trade is somewhat different from normal trade data, and there are issues with wholesale trade (discussed in Hillberry and Hummels 2003, and our online Appendix). While we need to be cautious in interpreting the border effects in the United States, it is not clear which way the bias goes relative to the European Union; indeed, it is not clear that the changes would be biased at all. Comparing EU15 countries in 1997 and 2017, as shown in Table 1, we see the revealed trade costs drop from 19 to 13 percent. Over the same period, the implied cost of trading across borders of American states falls from 11 to 10 percent. The confidence intervals on the ad valorem equivalents estimated for 2017 are sufficiently wide so that EU15 trade barriers are not significantly higher than US barriers. Moreover, if we consider the large EU28 (which includes the 2004 expansion into Eastern Europe), the point estimate of trade barriers actually falls below that for states in the United States.

When examining levels of border effects, differences in methods can make a big difference. For example, Head and Mayer (2000) estimate border coefficients ranging from 3.0 (1978–80) to 2.5 (1993–95). Using $\epsilon = 5$, this works out to ad valorem tariff equivalents of 82 and 65 percent. Using more comparable methods and data, we obtain in the online Appendix a border coefficient of 1.94 with a corresponding ad valorem equivalent of 47 percent in 1997 (with those methods, 28 percent for the states of the US economy). Another example of how border effects depend on the estimation method comes from Santamaría, Ventura, and Yesilbayraktar (2020). Taking advantage of newly available subnational flow data in the European Union

Table 1
EU Border Effects Compared to Their US Counterparts

Type of flow	United States		European Union		
	1997	2017	EU15 1997	EU15 2017	EU28 2017
Goods	11	10	19	13	8
Migrants	233	256	2,302	2,304	1,929
Mergers and Acquisitions	23	48	42	8	36

Note: Amount in each cell is the ad valorem equivalent (AVE) of τ for state or national borders. For migrants, the early year is 1995 (European Union) and 2000 (United States) and the late year is 2015. For mergers and acquisitions, the early period pools 1995–1998 and the later period pools 2015–2018.

(resembling the US Commodity Flow Survey), their approach does not require estimates of distance to self. In their study, the average ratio of within-border to cross-border log normalized market shares for comparable region pairs is 1.74, which implies a tariff equivalent of 42 percent for the period 2011–2017. While higher than our estimates, there are so many differences in data and method that we should use caution in comparing these amounts.

The second row of Table 1 compares border costs implied by within-EU migration to those implied by US state-to-state migration data. A first striking takeaway is how large those revealed tax rates are: leaving your European country of birth amounts to a tax-equivalent of 2300 percent (which barely changes over time, consistent with the flat migration costs since 1990 in Figure 2). The implied migration costs for the United States are also very high compared to trade costs, but an order of magnitude lower than the EU migration frictions. Mobility across American states is evidently easier than across EU countries, but still surprisingly costly.

How can we reconcile the de jure freedom of movement in the European Union since 1970 with these extremely high tax equivalents of migration frictions? One set of potential answers lies in the variety of institutions that erode the earnings or consumption potential for migrants. For example, for the purposes of pension benefits, years of work in one country are not always portable to the other countries. There are limitations on the transferability of employment insurance benefits. Finally, Baldwin and Wyplosz (2019) point out that recognition of professional qualifications remains imperfect: they give the example of French licensing rules that effectively exclude hairdressers from the rest of Europe. The United States also features some limitations of this kind, such as the requirement to pass legal exams in each state where a lawyer would like to practice.

The above examples notwithstanding, the estimated tax equivalents may be overestimated because the migration literature has underestimated the migration cost elasticity. Using instead the $\epsilon = 5$ that we do for trade flows, the ad valorem equivalent falls to 165 percent in the EU28 in 2017 and 51 percent for the United States. The trade literature has the advantage of using bilateral variation in trade costs, coming from tariffs or freight rates to estimate the cost elasticity. In contrast, the migration elasticity estimates rely on destination-level variables, such as real wages

or income taxes. When Eaton and Kortum (2002) use wage variation to estimate the elasticity for trade, they obtain a lower estimate than is typically found using tariff variation. We conjecture that if the migration gravity literature could find bilateral cost shifters, the elasticity estimates would be larger, lowering the implied tax of the border. Resolving this issue is also important because of the welfare implications of freer migration that are implied by this elasticity.

The final row of Table 1 shows the ad valorem equivalent of the implied border costs for merger and acquisition transactions in the United States and different definitions of the European Union. The US estimate of 48 percent for the 2015–2018 pooled years is surprisingly high. The European Union appears to exhibits lower cross-border frictions in the firm acquisition market, particularly when we confine the sample to the longer-standing EU15. From the late 1990s to the late 2010s, it appears that the EU15 market for corporate control has become considerably more free. One caveat is that even with the help of pooling four years in each period, the merger and acquisition data are noisy and the standard errors on the border effects are large. The confidence intervals for the various estimates of the ad valorem equivalent overlap considerably. The key point, however, is that for this measure of freedom of capital movement, the European Union is not significantly behind the United States and might even have passed it.

Kalemli-Özcan et al. (2010) and Ekinci, Kalemli-Özcan, and Sørensen (2009) examine capital market integration in the United States and the European Union from a very different angle and render a different verdict. While we focus on merger and acquisition flows because of the geographic detail, they use a broad measure of capital flows: the ratio of output (GDP) to income (GNI). This ratio diverges from one when the economy becomes a net recipient of income earned abroad. Kalemli-Özcan et al. (2010) find that their model-based test does not reject full capital market integration for states in the United States. Ekinci, Kalemli-Özcan, and Sørensen (2009) find less integration in the European Union and attribute some of the frictions to absence of trust between countries. These studies not only differ from our approach in terms of methodology, they also use earlier data; the US and EU data finish in 2000 and 2003, respectively.

The cross-sectional evidence on the levels of border effects in the United States and European Union should be viewed cautiously given the various measurement issues. However, the results provide additional evidence that the European integration project has succeeded in the market for goods, but less progress has been made on integration with respect to migration.

Price-Based Measurement of the EU Effect

The natural complement to measuring EU economic integration based on quantities in the context of a gravity model is to look at differences in prices. This approach is motivated by the Law of One Price: essentially, if the same good is being sold in the same market, it should sell for a single price. An implication is that as trade costs fall, competitive pressures should lead to convergence in the prices paid

by consumers. Anderson and van Wincoop (2004) use simulations from a modified version of Eaton and Kortum (2002) to show that the bilateral price gaps are informative about the extent of trade costs between two countries.

Price-based methods have mainly been applied to assess market integration for goods, but one can also view real wage gaps as measures of the degree of labor market integration. The incentive to migrate depends importantly on variation in real per capita income. Across a region where migration is relatively easy, we expect movement of people to put pressures on incomes to fall where they start out high and vice versa. This argument suggests that dispersion of real incomes across countries is complementary to the gravity evidence on migration: as border impediments decline, we should see both a decrease in the border effect in migration and shrinking dispersion of real incomes.

Price-Level Index Convergence

We start with aggregate evidence on price convergence, taking advantage of an OECD dataset on purchasing power parities (PPP) for each country going back to 1960. The price level index of GDP is calculated as the PPP exchange rate per US dollar divided by the market exchange rate in each year. We take logs and calculate the standard deviation year by year over different subsets of countries and years.

The number of EU members has expanded over time, bringing dissimilar countries into the Union. When measuring price-level convergence, it is thus useful to look at different groups. In Figure 5, the heavy purple line shows price-level index convergence for the original six members of the Treaty of Rome in 1957. The light blue line shows price convergence across the 15 countries that joined the European Union by 1995. The light black line shows price convergence across the 28-country membership of the European Union, including the countries that joined after 2004, but before the exit of the United Kingdom.

While the solid lines in Figure 5 show different groups of countries; the vertical dashed lines show dates when additional countries joined the European Union. Figure 5 reveals three periods of price convergence. The first is from 1960 to 1971, which coincides with the removal of tariffs and quantitative restrictions amongst the original six members (shown in purple in the figure). The broader group of 15 countries who join the European Union by 1995 does not exhibit this convergence at this time, suggesting that the price-level convergence among the EU6 was not some general European effect like reduced transportation costs. For the EU15 group of countries (the light blue line), a clear phase of convergence follows the accession of low-price countries Portugal and Spain in 1986. Finally there is some modest convergence of the full set of EU members following the 10-country accession in 2004. The two cases of divergence in the EU6 price levels in the late 1970s and early 1990s come from real depreciation of the Italian lira relative to the other original members. From 1995 forward, Italy's prices cease to deviate markedly from the other core EU countries (for an illustration of price level convergence in Italy, see online Appendix figure B.1).

How does this price convergence across Europe compare to the experience of the United States? Starting in 2008, the US government began to estimate what it

Price-Level Variation across EU Members and US States
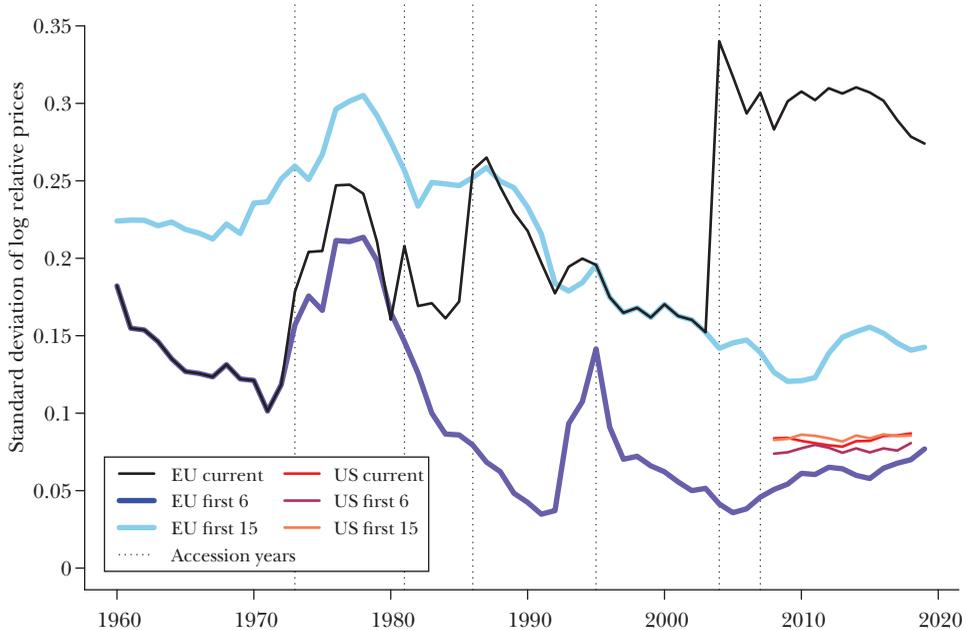
Source: Source: See online Appendix B.6 for data sources.
Note: The first 6/15 US states are determined by order of admission to the European Union. For the EU, the first 6 are signatories of the Treaty of Rome (1957) and the first 15 are members as of 1995.

calls regional price parities. To preserve an analogy with the way we carry out the EU calculations, we consider the whole set of 50 states as well as the first six and the first 15 American states. The latter subsets comprise more proximate economies. The striking finding is that original EU founders have slightly less price-level dispersion than American states, regardless of which set of states we use. However, the broader EU28 group of countries exhibits more than three times as much price dispersion as American states, pointing to incomplete integration of the more recent joiners.

Measuring overall price-level convergence has the advantage of comprehensiveness: both goods and services enter the index, weighted according to their relative importance in the economy. The disadvantage is that most services are not actively traded across borders. Furthermore, there is little reason to believe that the mix of product varieties and qualities offered in say, Portugal, are the same as those available in Denmark. These considerations motivate us to investigate whether prices of specific, consistently defined, tradeable goods have been converging over time.

Product-Level Price Convergence: A Case Study of Three Cars

Several studies have examined the relationship between EU integration and price convergence at the level of detailed products, typically using a sample of products tracked over time. Like us, Crucini, Telmer, and Zachariadis (2005) measure

price dispersion as the standard deviation of log prices. Their sample has 594 products in 1975, and this figure rises to 1,101 in 1990. Furthermore, their survey adds the low-wage countries Greece, Portugal, and Spain in 1980. These changes may explain why price dispersion jumps by 6 percentage points from 1975 to 1980. However, price gaps in their study then fell slightly from 0.28 in 1980 to 0.26 in 1990. Changes in the relative importance of goods might matter, as the authors show that there is a lot of heterogeneity across goods. This points to the importance of keeping the sample of products and countries constant.

Méjean and Schwellnus (2009) also analyze cross-country prices for detailed (eight-digit) products. They consider exports from France to the EU15 countries (those that had joined the European Union by 1995) and to the rest of the world. They show lower absolute price variation within the European Union, but it is fairly flat at about 30 percent from 1995 to 2004. They interpret their results as market integration increasing arbitrage pressures. Verboven et al. (2020) examine price dynamics for desktops, laptops, smartphones, and tablets for ten EU members from 2012 to 2015. They find "international price differences appear to be large and persistent . . . . There is no obvious indication that price differences are lower online." Given the short time period, their study is silent on whether cross-sectional variation has been declining over time.

Large gaps in car prices across EU members have long been the subject of scrutiny. In response, the European Commission (2011) reported price data for a large set of car models between 1993 and 2011. They collected prices on versions of each model with common engine size and trim across all countries. We augmented the 1993–2011 European Community dataset with Goldberg and Verboven (2005) data from 1970 to 1999 for Germany, France, Italy, the United Kingdom, and Belgium, giving us four decades of data on a stable set of products. To preserve a standard comparison over the longest possible period, we focus on three major models that have been available in all major EU markets since the 1970s: the Honda Civic (introduced in 1973), the VW Golf (1974), and the Ford Fiesta (1976).

Figure 6 plots the evolution of the standard deviation (across countries) of the log of these three car prices across countries, on a before-tax basis. The black squares show price dispersion in the five large markets from 1973 to 2011 using both datasets. The blue circles depict standard deviations over the 15 members of the European Union since 1995 (relying solely on the European Commission data). Although we use just three models, they do not appear to be outliers. Dvir and Strasser (2018) use all models from the European commission sample and also calculate standard deviations of the log pre-tax price fluctuating in the 0.05–0.1 range over the 1993–2011 period.

The main takeaway of Figure 6 is that after a period of noisy dispersion of car prices in the 1970s, we mainly see convergence in prices in the 1980s and 1990s for the five major markets shown with black squares. Degryse and Verboven (2000) review the large hedonic price literature on car prices and report similar time patterns. The declining trend in price dispersion starting in the late 1980s is also consistent with Goldberg and Verboven (2005), who find strong support for the hypothesis that European integration resulted in price convergence. In the late

Figure 6
Car Price Differentials
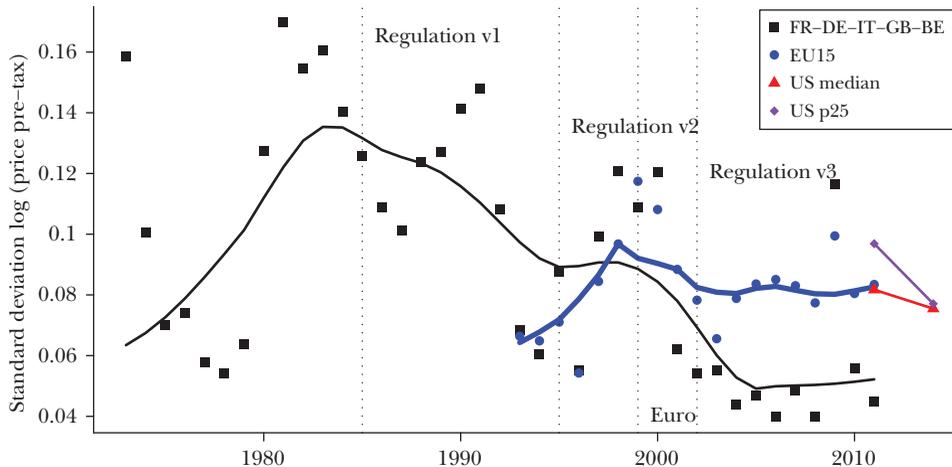
Source: See online Appendix B.6 for data sources.
Note: Regulations v1, v2, and v3 refer to the three versions of EU texts regulating the block exemption of the motor vehicles industry regarding distribution (see https://ec.europa.eu/competition/sectors/motor_vehicles/legislation/legislation_archive.html).

1990s, the core five countries shown by the black squares and the broader EU15 have similar levels of dispersion. A period of price convergence follows that is more pronounced for the core group of five countries than for the EU15 countries.[7]

While much of the price dispersion seems related to exchange rate fluctuations, actions by the European Commission also appear to have curbed the tendency of firms to segment the market and price discriminate across consumers. Vertical dotted lines in Figure 6 shows the timing of revisions of the rules about how carmakers can choose their distribution system. Our figures (including the ones in Appendix H) show that the 2002 revision might have been the most efficient in promoting car price convergence. In 2011, when the European Commission discontinued reporting of model-level prices for cars (at https://ec.europa.eu/competition/sectors/motor_vehicles/prices/report.html), it explained, "the situation has improved greatly, in part due to enforcement action by the Commission, and also thanks to the increased availability of price information on the internet."

As a way of benchmarking the level of price convergence for these three models of cars, we compare the level of price dispersion across US states. For that purpose, we use Consumer Expenditure Survey micro data from 2010 to 2015. When

---

[7]Online Appendix H.3 shows the same figure including taxes. This does not change the pattern much for the five nations shown by the black squares in Figure 6, but it leads to much wider dispersion in prices in the EU15 countries, driven chiefly by the high taxes charged in Denmark and Finland. Also, Appendix H shows the shorter sample constructed by the European Commission that covers a wider range of countries. This data shows greater convergence in the EU15 countries, a pattern also reported by Gil-Pareja and Sosvilla-Rivero (2008) and Sosvilla-Rivero and Gil-Pareja (2012).

disaggregated to the state level, the number of observations for small-population states is sometimes too small to be reliable. Therefore, we pool the state-level data into two periods, 2010 to 2012 and 2013 to 2015 and compare those two periods (for further details, see online Appendix B.6). Since the Consumer Expenditure Survey does not provide model-level prices, we compute measures of dispersion based on the median and 25th percentile prices. The latter may be more appropriate since the Fiesta, Civic, and Golf are relatively inexpensive cars, but it turns out not to matter much. The bottom line is that car price dispersion across the EU15 seems to have reached a level strikingly close to the one observed in the United States. Aside from 2009 (when UK pound depreciation increased dispersion), the five large EU countries appear to have less car price dispersion (0.04–0.06) than the US states (0.08–0.11). The efforts of the European Commission to reduce car price dispersion appear to have been successful in the five large EU markets.

As a final comparison, we use state-level incomes in the United States and national-level incomes in Europe, adjusted by purchasing power parity exchange rates, to calculate the standard deviations of real log incomes in each period. For the United States, we use state-level personal incomes from the US Bureau of Economic Analysis adjusted by the corresponding regional price parities. For EU real incomes, we combined data from the Penn World Tables and World Development Indicators (online Appendix B.6 provides detail on the calculations).

It is again useful to divide the results for Europe into the original six countries that signed the Treaty of Rome in 1957, the EU15 group that were members by 1995, and the EU28 group that represents the countries that are currently members of the European Union (although this data includes the United Kingdom, which has of course recently left the Union). Figure 7 shows that adding new members creates big jumps in real wage variation, followed by periods of steady convergence. This occurs after the admission of Ireland in 1973, Portugal and Spain in 1986, and even more impressively after the eastern EU enlargements in 2004 and 2007.[8] Looking at fixed sets of countries eliminates the upward jumps that occur when low-wage countries join. The founding EU members have a downward trend in real income variation that lasted until the early 1980s. The real incomes of the 15 countries who had joined by 1995 mainly converged until the most recent decade. The most remarkable finding in Figure 7 is that even after a recent resurgence of inequality, dispersion among the founding six EU nations is about the same as among the US states: about 0.10 to 0.13 over the last decade.

Conclusion

In terms of formal institutions, the European Union is not on the verge of becoming a "United States of Europe." But on multiple fronts, EU economic

---

[8]Various studies have considered the evidence on real wage or income convergence in the European Union, but most have focused on "beta convergence" and unit root tests. Quah (1993) compellingly argues that this approach suffers from Galton's fallacy.

Figure 7
Real Income Variation across EU Members and US States
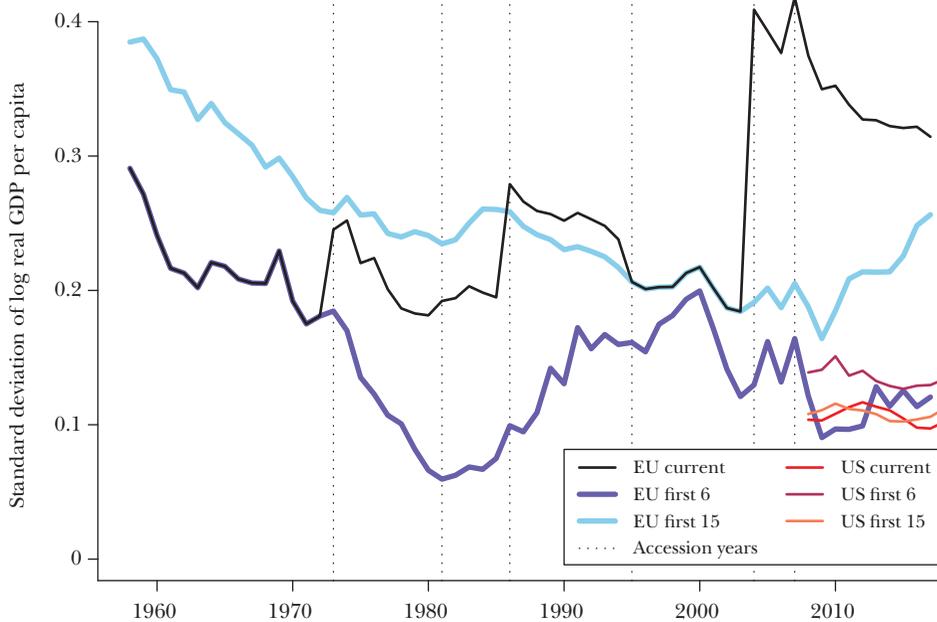
Source: See online Appendix B.6 for data sources.
Note: The first 6/15 US states are determined by order of admission to the Union. For the EU, the first 6 are signatories of the Treaty of Rome (1957) and the first 15 are members as of 1995.

integration now matches or even beats the equivalent measure for the 50 American states. This is remarkable. The United States has more than 230 years as a federal state with a constitutional prohibition on barriers to interstate commerce. Of course, all comparisons with the United States require caution since the last two additions for the United States occurred in 1959, whereas 22 countries have joined the European Union since that year, with the most recent entrant (Croatia) joining in 2013. Perhaps the most useful comparison across the US states is with the EU15, which includes the entry of some lower-income states but has had constant membership since 1995. The border tax equivalents implied by flows of goods and merger and acquisition transactions within the EU15 have reached the levels estimated for US states. When measuring integration as convergence in price levels, the EU15 is quite similar to the American states. Focusing on a product for which we have detailed and comparable measures across all countries—compact, mass-market cars—we confirm the finding for the aggregate price index: the American states and the EU15 are again very similar.

Regarding what may be the most politically sensitive of the four movements, migration, our estimates suggest that barriers remain considerably higher in Europe. Despite the absence of formal restrictions on movement, Europeans act as if their

human capital is very heavily taxed by moving countries. This lack of mobility across European borders likely reflects a variety of labor market frictions and cultural differences. On the other hand, the incentives to move have fallen substantially within the EU6, with dispersion in real incomes now essentially the same as that in core eastern states of the United States. Real income variation is three times as high in the European Union as a whole, but enlargement has been followed by a trend towards equalization, so there is little reason to think the EU28 has reached a steady state in terms of income disparities across its members.

A potential objection to the rosy view of EU integration depicted above is that a rising inward orientation for the European Union could arise from construction of higher barriers imposed on flows from the rest of the world—unification via "Fortress Europe." However, adapting the gravity equation to allow for such a possibility, we find the opposite tendency prevails. With respect to goods, services, and migration, the European Union is increasingly open to the rest of the world. Rising intra-EU flow shares have come from falling intra-national shares, precisely the pattern needed for welfare to rise in standard trade models.

References

Ahern, Kenneth R., Daniele Daminelli, and Cesare Fracassi. 2015. "Lost in Translation? The Effect of Cultural Values on Mergers around the World." *Journal of Financial Economics* 117 (1): 165–89.

Alesina, Alberto, and Roberto Perotti. 2004. "The European Union: A Politically Incorrect View." *Journal of Economic Perspectives* 18 (4): 27–48.

Anderson, James E., and Eric van Wincoop. 2004. "Trade Costs." *Journal of Economic Literature* 42 (3): 691–751.

Arkolakis, Costas, Arnaud Costinot, and Andrés Rodríguez-Clare. 2012. "New Trade Models, Same Old Gains?" *American Economic Review* 102 (1): 94–130.

Baier, Scott L., and Jeffrey H. Bergstrand. 2007. "Do Free Trade Agreements Actually Increase Members' International Trade?" *Journal of International Economics* 71 (1): 72–95.

Baier, Scott L., Jeffrey H. Bergstrand, and Michael Feng. 2014. "Economic Integration Agreements and the Margins of International Trade." *Journal of International Economics* 93 (2): 339–50.

Baldwin, Richard, and Charles Wyplosz. 2019. *The Economics of European Integration.* 6th ed. New York: McGraw Hill.

Beine, Michel, Simone Bertoli, and Jesús Fernández-Huertas Moraga. 2016. "A Practitioners' Guide to Gravity Models of International Migration." *The World Economy* 39 (4): 496–512.

Bown, Chad P., and Douglas A. Irwin. 2016. "The GATT's Starting Point: Tariff Levels Circa 1947." The World Bank Policy Research Working Paper 7649.

Brunsden, Jim. 2019. "Brussels Acts against All EU Members over Single Market Failings." *Financial Times*, June 6. https://www.ft.com/content/d073f25a-885f-11e9-a028-86cea8523dc2.

Caliendo, Lorenzo, Fernando Parro, Luca David Opromolla, and Alessandro Sforza. 2020. "Goods and Factor Market Integration: A Quantitative Assessment of the EU Enlargement." https://sites.google.com/site/lorenzocaliendo/research/cops.

Churchill, Winston. 1946. "United States of Europe." Speech, University of Zurich, September 19, 1946. https://winstonchurchill.org/resources/speeches/1946-1963-elder-statesman/united-states-of-europe/.

Coeurdacier, Nicolas, Roberto A. De Santis, and Antonin Aviat. 2009. "Cross-border Mergers and Acquisitions and European Integration." *Economic Policy* 24 (57): 56–106.

Costinot, Arnaud, and Andrés Rodríguez-Clare. 2014. "Trade Theory with Numbers: Quantifying the Consequences of Globalization. In *Handbook of International Economics*, Vol. 4., edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 197–261. Amsterdam: Elsevier.

Crucini, Mario J., Chris I. Telmer, and Marios Zachariadis. 2005. "Understanding European Real Exchange Rates." *American Economic Review* 95 (3): 724–38.

Degryse, Hans, and Frank Verboven. 2000. *Car Price Differentials in the European Union: An Economic Analysis*. London: CEPR.

Dvir, Eyal, and Georg Strasser. 2018. "Does Marketing Widen Borders? Cross-Country Price Dispersion in the European Car Market." *Journal of International Economics* 112: 134–49.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, Geography, and Trade." *Econometrica* 70 (5): 1741–79.

Eaton, Jonathan., Samuel Kortum, and Sebastian Sotelo. 2013. "International Trade: Linking Micro and Macro." In *Advances in Economics and Econometrics Tenth World Congress: Applied Economics*, Vol. II, edited by Daron Acemoglu, Manuel Arellano, and Eddie Dekel. Cambridge, UK: Cambridge University Press.

Ekinci, Mehmet Fatih, Şebnem Kalemli-Özcan, and Bent E. Sørensen. 2009. *Financial Integration within EU Countries: The Role of Institutions, Confidence and Trust*. Chicago: University of Chicago Press.

European Commission. 1985. "Completing the Internal Market." White Paper COM (85) 310.

European Commission. 2011. Car price reports. https://ec.europa.eu/competition/sectors/motor_vehicles/prices/report.html.

Feldstein, Martin. 1997. "The Political Economy of the European Economic and Monetary Union: Political Sources of an Economic Liability." *Journal of Economic Perspectives* 11 (4): 23–42.

Gil-Pareja, Salvador, and Simón Sosvilla-Rivero. 2008. "Price Convergence in the European Car Market." *Applied Economics* 40 (2): 241–50.

Goldberg, Pinelopi K., and Frank Verboven. 2005. "Market Integration and Convergence to the Law of One Price: Evidence from the European Car Market." *Journal of International Economics* 65 (1): 49–73.

Grogger, Jeffrey, and Gordon H. Hanson. 2011. "Income Maximization and the Selection and Sorting of International Migrants." *Journal of Development Economics* 95 (1): 42–57.

Head, Keith, and Thierry Mayer. 2000. "Non-Europe: The Magnitude and Causes of Market Fragmentation in the EU." *Review of World Economics* 136 (2): 284–314.

Head, Keith, and Thierry Mayer. 2004. "Market Potential and the Location of Japanese Investment in the European Union." *Review of Economics and Statistics* 86 (4): 959–72.

Head, Keith, and Thierry Mayer. 2010. "Illusory Border Effects: Distance Mismeasurement Inflates Estimates of Home Bias in Trade." In *The Gravity Model in International Trade: Advances and Applications*, edited by Peter A.G. van Bergeijk and Steven Brakman, 165–192. Cambridge, UK: Cambridge University Press.

Head, Keith, and Thierry Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." In *Handbook of International Economics*, Vol. 4, edited by Gita Gopinanth, Elhalan Helpman, and Kenneth Rogoff, 131–196. Amsterdam: Elsevier.

Head, Keith, Thierry Mayer, and John Ries. 2009. "How Remote Is the Offshoring Threat?" *European Economic Review* 53 (4): 429–44.

Head, Keith, and John Ries. 2001. "Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of US-Canada Trade." *American Economic Review* 91 (4): 858–76.

Head, Keith, and John Ries. 2008. "FDI as an Outcome of the Market for Corporate Control: Theory and Evidence." *Journal of International Economics* 74 (1): 2–20.

Heiland, Inga, Gabriel Felbermayr, and Jasmin Groeschl. 2020. "Complex Europe: Quantifying the Cost

of Disintegration." CEPR Discussion Paper 15200.

Hillberry, Russell, and David Hummels. 2003. "Intranational Home Bias: Some Explanations." *Review of Economics and Statistics* 85 (4): 1089–92.

Hugo, Victor. 1849. Speech to the Peace Congress of Paris. https://www.thenewfederalist.eu/170-years-since-victor-hugo-s-speech-about-the-united-states-of-europe? lang=fr.

Kalemli-Özcan, Şebnem, Ariell Reshef, Bent E. Sørensen, and Oved Yosha. 2010. "Why Does Capital Flow to Rich States?" *The Review of Economics and Statistics* 92 (4): 769–83.

Kleven, Henrik, Camille Landais, Mathilde Muñoz, and Stefanie Stantcheva. 2020. "Taxation and Migration: Evidence and Policy Implications." *Journal of Economic Perspectives* 34 (2): 119–42.

Limão, N. 2016. "Preferential Trade Agreements." In *Handbook of Commercial Policy*, Vol. 1B, edited by Kyle Bagwell and Robert W. Staiger, 281–367. Amsterdam: Elsevier.

Mayer, Thierry, Vincent Vicard, and Soledad Zignago. 2019. "The Cost of Non-Europe, Revisited." *Economic Policy* 34 (98): 145–99.

Méjean, Isabelle, and Cyrille Schwellnus. 2009. "Price Convergence in the European Union: Within Firms or Composition of Firms?" *Journal of International Economics* 78 (1): 1–10.

Novy, Dennis. 2013. "Gravity Redux: Measuring International Trade Costs with Panel Data." *Economic Inquiry* 51 (1): 101–21.

Palmer, Glenn, Vito D'Orazio, Michael Kenwick, and Matthew Lane. 2015. "The MID4 Dataset, 2002–2010: Procedures, Coding Rules and Description." *Conflict Management and Peace Science* 32 (2): 222–42.

Quah, Danny. 1993. "Galton's Fallacy and Tests of the Convergence Hypothesis." *The Scandinavian Journal of Economics* 95 (4): 427–43.

Redding, Stephen J. 2016. "Goods Trade, Factor Mobility and Welfare." *Journal of International Economics* 101: 148–167.

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

Santamaría, Marta, Jaume Ventura, and Ugur Yesilbayraktar. 2020. "Borders within Europe." Barcelona Graduate School of Economics Working paper 1229.

Schuman, Robert. 1950. "The Schuman Declaration." Speech, May 9, 1950. https://europa.eu/european-union/about-eu/symbols/europe-day/schuman-declaration_en.

Sosvilla-Rivero, Simón, and Salvador Gil-Pareja. 2012. "Convergence in Car Prices among European Countries. *Applied Economics* 44 (25): 3247–54.

Tabuchi, Takatoshi, and Jacques-François Thisse. 2002. "Taste Heterogeneity, Labor Mobility and Economic Geography." *Journal of Development Economics* 69 (1): 155–77.

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

Verboven, Frank, Nestor Duch-Browne, André Romahn, and Lukasz Grzybowski. 2020. "Are Online Markets More Integrated than Traditional Markets? Evidence from Consumer Electronics." CEPR Discussion Paper 15031.

Weidner, Martin, and Thomas Zylkin. 2019. "Bias and Consistency in Three-Way Gravity Models." Unpublished.

Migration and Labor Market Integration in Europe

David Dorn and Josef Zweimüller

The Treaty of Rome, signed in 1957 by Belgium, France, Italy, Luxembourg, the Netherlands, and West Germany, envisioned the development of a common market with free movement of goods, capital, services and persons. Today, legal barriers to labor mobility across European countries have been dismantled: more than 460 million citizens of 31 European countries can choose to reside in any other partner country, they can work there without needing a work permit, and they are entitled to equal treatment with nationals in access to employment and public services.

However, the European labor market remains considerably less integrated and more heterogeneous than the US labor market, which comprises a population of 330 million across the 50 states. For example, consider the dispersion of unemployment rates. In 2019, national unemployment rates in European countries were as low as 2.0 percent in Czechia and 3.2 percent in Germany, but as high as 13.7 percent in Spain and 16.6 percent in Greece (Eurostat 2020a). By comparison, state-level unemployment rates within the United States ranged from 2.4 percent to 6.1 percent (Bureau of Labor Statistics 2019). The European labor market also has much lower levels of spatial mobility. The share of European citizens living in a different country than their country of birth was less than 5 percent in 2019

■ *David Dorn is UBS Foundation Professor of Globalization and Labor Markets and Josef Zweimüller is Professor of Macroeconomics and Labor Markets, both at the University of Zurich, Zurich, Switzerland. Both authors are Research Fellows, Centre for Economic Policy Research, London, United Kingdom. Dorn is the corresponding author at david.dorn@econ. uzh.ch.*

(Eurostat 2020b), while the fraction of cross-state migrants in the US population has long been close to one-third (Molloy, Smith, and Wozniak 2011). However, whereas domestic mobility in the United States appears to be slowly declining, it is increasing in the European labor market.

The removal of restrictions to international migration has very large potential economic benefits, which may exceed the benefits of other integration measures such as free trade by an order of magnitude (Clemens 2011; Dustmann and Preston 2019). Nonetheless, European labor market integration remains a contentious policy issue. Skepticism about immigration is a signature issue of right-wing European populism (Margalit 2019; Guiso et al. 2020) and is also strongly correlated with general distrust towards the European Union (Jeannet 2017).

In this article, we discuss the past, present, and potential future of the European labor market. We begin by documenting patterns of labor mobility across European countries. We next ask whether and to what extent the labor markets of these countries have become more integrated over time. Finally, we discuss remaining obstacles for European labor market integration. Our primary focus is on migration between European countries: for surveys of the literature on overall immigration in Europe, useful starting points are Dustmann and Frattini (2011) and De La Rica, Glitz, and Ortega (2015).

Labor Market Integration in Europe

In the Treaty of Rome, the six founding members of the European Economic Community agreed on the free movement of citizens within those countries, thus extending to the entire economy the labor-mobility agreement for the coal and steel industries that had been introduced by the 1951 Treaty of Paris. The Schengen Agreement of 1985 further led to the fall of national border controls, which facilitated cross-border work.[1] However, despite a harmonization of visa policies, each country maintained the right to apply its own rules for the provision of work visas to citizens of countries that do not participate in the common market.

In successive enlargements, six additional Western European countries joined the European Economic Community (EEC): the United Kingdom, Ireland, and Denmark in 1973, Greece in 1981, and Portugal and Spain in 1986. In addition, East Germany was integrated into the bloc following the German unification in 1990. The 1992 Treaty of Maastricht established the European Union (EU), whose goal was a closer political integration among the EEC members, including the establishment of EU citizenship. In the same year, the twelve members of the EEC and the seven members of the European Free Trade Association (EFTA) signed an

---

[1]All but six of the countries that eventually participated in the common European labor market, also became part of the Schengen area. The United Kingdom and Ireland opted out of joining the Schengen agreement, while some of the newest members of the European market in southeastern Europe are obliged to join in the future.

agreement to expand the common market beyond the EEC/EU by forming the European Economic Area (EEA), which covered nearly all of Western Europe. The EU and EEA then expanded eastwards and added a further 13 countries from 2004 onwards.[2]

Thus, since its foundation, the common European Economic Area labor market grew from six countries with a population of 167 million in 1957 to 32 countries with a population of about 530 million in 2020. We will refer to these countries as "EEA countries," and include Switzerland in that group, which participates in the common market despite not being an EEA member, and the United Kingdom, which left the common market in 2021. Prior to "Brexit," which reduced the expanse of the European labor market for the first time, the common market included all countries on the European continent, except most of the successor countries of the USSR and of Yugoslavia, as well as Turkey, Albania, and some micro-states.

The changing membership in the European Economic Area had large implications for the dispersion of the material standard of living among member states. Figure 1 ranks the EEA population by the real per capita income of their country of residence and shows the difference between the EEA resident at the 5th versus the 95th percentile: In 1958, an Italian resident was at the 5th and a German at the 95th percentile; in 2016, a Romanian was at the 5th while someone from Holland was at the 95th percentile.[3] The figure indicates that per capita income differentials—indicated by the 95/5 percentile ratio—have increased over time. The 95/5 percentile ratio expanded from 1.19 to 1.85 with the accession of Spain and Portugal in 1986, and later jumped to 2.82 and 3.47 following the eastern enlargements of 2004 and 2007, respectively. For comparison, the 95/5 percentile ratio among US states was 1.86 in 2018, with New York and South Carolina being the states at the 95th and 5th percentiles, respectively.

The eastern enlargement of the European Union and the resulting sharp increase in income differentials within the common labor market created the potential for substantial migration from poorer to richer countries. Most older

---

[2]Depending on data availability, we will subsequently report statistics for the following country groups: "EU-15" comprises the twelve European countries that had already been members of the EEC by 1986 (Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, United Kingdom), plus three countries that joined the EEA in 1994 as members of EFTA and subsequently acceded to the European Union in 1995 (Austria, Finland, Sweden); "EFTA" comprises three EFTA members which joined the EEA in 1994 or 1995 (Iceland, Liechtenstein, and Norway), and one that rejected an accession to the EEA in a referendum but later joined the common market via bilateral treaties in 2005 (Switzerland); "EU-28" comprises the EU-15 plus 13 countries that joined the European Union in 2004 (Cyprus, Czechia, Estonia, Hungary, Latvia, Lithuania, Malta, Poland, Slovakia, Slovenia), in 2007 (Bulgaria and Romania) or in 2013 (Croatia); and "EU-27" is the same set of countries minus the United Kingdom following its exit from the European Union in 2021.

[3]When Greece joined the European Union in 1981, it became the poorest country among existing member states, but its population comprised less than 5 percent of EU residents. In 1986, Spain replaced Italy at the 5th percentile position, before that spot was taken over by Greece in 1990. From 2004 onwards, several Eastern European countries (Poland, Bulgaria, and Romania) held the 5th percentile spot. The 95th percentile position was usually held by Germany or the Netherlands, with brief interruptions by Austria, Ireland, Sweden, and Denmark.

Figure 1
Income Dispersion among Countries in the Common European Labor Market
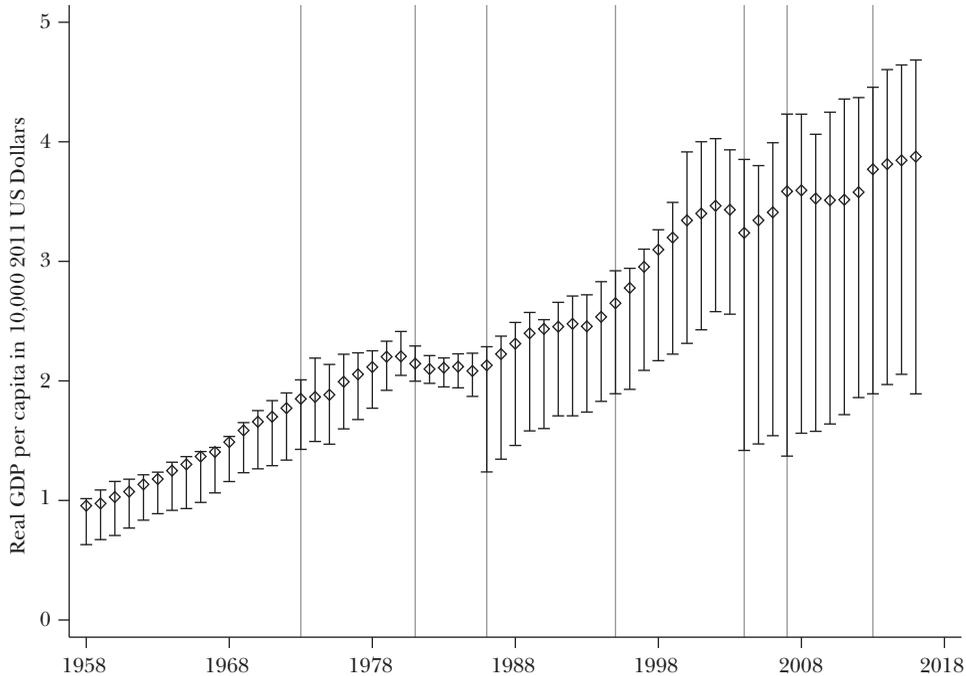
Source: Author's calculation based on data from Bolt and van Zanden (2020).
Note: The figure shows the distribution of real GDP per capita across countries that in a given year were part of the common European labor market through membership in what was the European Economic Community and has evolved into the European Economic Area. Diamonds indicate the median value of the population-weighted distribution, while whiskers indicate the range between the 5th and 95th percentile. Vertical lines mark years in which new countries joined the European Union.

member states, led by Germany and Austria who are in close proximity to the new Eastern European entrant countries, initially imposed rules that restricted the access of workers from new members states to their labor markets for a transitional period of up to seven years (Fihel et al. 2015). Only the United Kingdom, Ireland, and Sweden immediately opened up their labor markets in 2004, and these countries received large inflows of Eastern European citizens as a consequence.

Patterns of Migration

To document the extent of migration within Europe over longer time periods, we first look at Germany, the largest and wealthiest of the six EEC countries. We then take a closer look at migration patterns all over Europe since the year 2000

Figure 2
Annual Inflows of Foreign Citizens into Germany
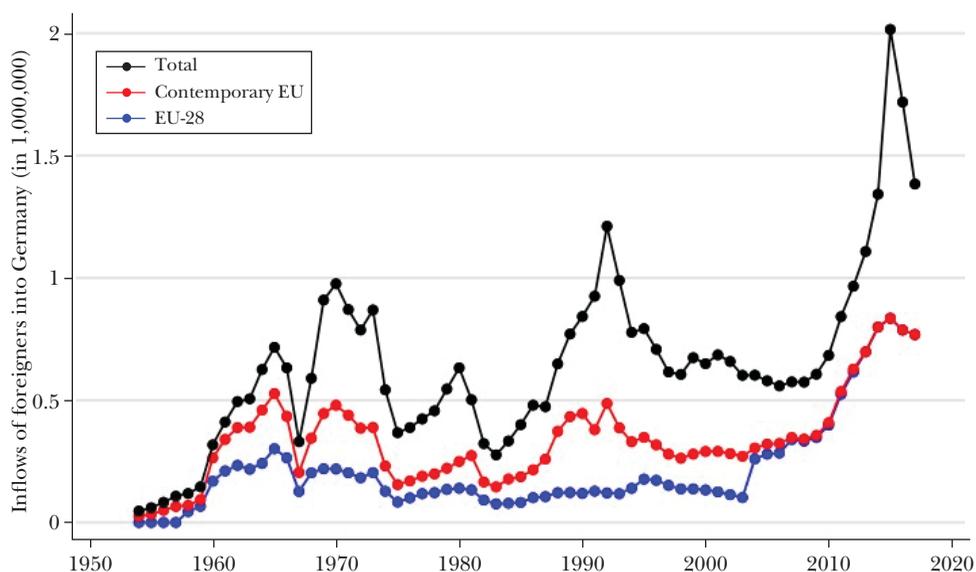
Source: International Migration Institute (2015), Statistisches Bundesamt (2020a)
Note: "Contemporary EU" indicates inflows of foreign nationals who were citizens of a country that was a member of the EEC/EU in the indicated year. "EU-28" indicates inflows of foreigners who were citizens of one of the 28 countries that eventually joined the European Union.

when within-EU migration flows started to surge. Finally we show that since 2000, within-EU migration has become increasingly high-skilled.

Immigration into Germany since the 1960s

Figure 2 depicts the annual inflow of immigrants into Germany since the mid-1950s. It indicates that sizable immigration commenced in the early 1960s, when "guest workers" were attracted to fill labor shortages in the booming "*Wirtschafts-wunder*" economy. During the 1960s and early 1970s, Germany was the leading destination of migrants from within the European Economic Community, while Italy was the most important origin country of these within-EEC migrants (Straub-haar 1988). However, many immigrants to Germany during the 1960s and 1970s originated from countries that were yet to join the EEC, such as Spain. From the 1970s onwards, a large fraction of immigrants also came from countries that never became part of the European labor market, such as Turkey. Immigration within the European Economic Area increased rapidly only after 2011, when the citizens of the Eastern European countries that had joined the European Union in 2004 gained full access to the German labor market following the expiry of the seven-year

transitional arrangements. In 2015, a continuously high inflow of migrants from Eastern Europe and a wave of refugees, primarily from Syria, led to a record immigration of two million individuals in a single year.[4]

Migration to and within Europe since the 2000s

The fact that inter-European migration increased strongly in recent years, as shown for the German case in Figure 2, motivates us to take a closer look at the last two decades. Column 1 of Table 1 reports the fraction of foreign nationals in the 2019 population of each country in the European Economic Area (except Liechtenstein), with countries listed in descending order of their 2015 GDP per capita. The second and third columns separate this total into the proportion of EU-27 and non-EU-27 foreigners in a country. The fourth column indicates the number of a country's citizens that reside in another EEA country, expressed as a percentage of the source country's domestic population.[5]

The table illustrates that immigrant stocks are positively correlated with countries' income levels: for example, the share of foreign nationals in the domestic population is largest in Luxembourg (47.5 percent) and Switzerland (25.1 percent), which are among the countries with highest incomes per capita worldwide. Many of the poorer Eastern European members have small foreigner shares in their populations, such as 0.6 percent in Romania and 0.8 percent in Poland. The contrast becomes stronger still if one focuses only on foreigners with EU citizenship in column 2 of Table 1. The destinations of international migrants within the EEA are almost entirely the higher-income countries of Western Europe. Instead, most of the foreign citizens living in the eastern countries of the European Economic Area come from non-EU nations, such as Russians residing in Estonia or Bosnians in Slovenia.

The patterns for emigrants, shown in column 4 of Table 1, are opposite to those for immigrants. Emigrants from Eastern Europe account for a large portion of citizens living in a different EEA country. Most strikingly, roughly one of every five Romanian citizens in the European Economic Area—a total of 3.6 million individuals—is living outside of Romania. Some of the southern member states, like Portugal or Greece, also have large diasporas elsewhere in Europe. By contrast, wealthier countries of Western Europe, like Germany or the United Kingdom, have relatively few of their citizens living abroad, at least compared to the much larger number of EEA and non-EEA foreigners that these countries host.

---

[4] Historically consistent time series for migration inflows are available for Germany and the Netherlands, but not for the other two largest founding members of the European Economic Community: France and Italy. Online Appendix Figure A1 shows time-series data on immigration to the Netherlands, which are similar to those for immigration to Germany.

[5] There are more comprehensive European migration statistics based on individuals' nationality rather than their country of birth. In 2018, 86 percent of the foreign nationals residing in EU-15/EFTA countries were born abroad. That fraction is lower in most Eastern member states, and as low as 49 percent in Bulgaria and Lithuania (Eurostat 2020). Data on a country's emigrants is available only for those who reside in European Economic Area countries, but not for those who moved to a non-EU country.

Table 1

Foreign Citizens Residing in EEA Countries in 2019 and Change in Foreign Citizens Residing in EEA Countries, 2004 to 2019

	Foreign citizens living in a country, in percent of country's population, 2019			*Country's citizens living in other EU country, in per-cent of country's population, 2019*	*Change %pts of foreign citizens living in a country, 2004–2019*	
	All foreign nationalities	*EU nationalities*	*non-EU nationalities*		*EU nationalities*	*non-EU nationalities*
	(1)	(2)	(3)	(4)	(5)	(6)
I. Countries with per capita income above EU average						
Luxembourg	47.5	40.1	7.4	5.5	5.8	1.9
Switzerland	25.1	16.5	8.6	1.1	4.8	−0.1
Ireland	12.5	9.2	3.3	8.4	1.0	−0.1
Norway	11.0	6.8	4.2	1.6	4.3	1.5
Iceland	12.4	10.5	1.9	7.4	4.9	0.1
Denmark	9.1	3.9	5.2	2.4	2.6	1.4
Netherlands	6.4	3.3	3.1	3.4	1.9	0.2
Sweden	9.1	3.1	6.0	2.0	0.8	3.0
Austria	16.2	8.2	8.0	3.0	5.7	1.3
Finland	4.7	1.8	2.9	2.5	1.1	1.5
Germany	12.2	5.3	6.9	1.5	2.3	1.0
Belgium	12.3	8.0	4.3	2.5	2.3	1.7
United Kingdom	9.3	5.5	3.8	1.4	3.5	0.6
France	7.3	2.4	4.9	1.4	0.4	1.1
Italy	8.7	2.6	6.1	3.4	2.0	3.3
Malta	16.9	9.2	7.7	2.5	6.8	6.0
Spain	10.3	4.2	6.1	1.7	2.1	1.4
II. Countries with per capita income below EU average						
Cyprus	17.8	13.4	4.3	3.9	6.2	0.1
Slovenia	6.6	1.0	5.6	3.6	0.9	3.5
Estonia	15.1	1.6	13.5	7.0	1.1	−3.6
Czechia	5.2	2.2	3.1	1.6	1.5	1.8
Portugal	4.7	1.5	3.1	14.3	0.6	−0.1
Lithuania	1.7	0.3	1.4	15.8	0.2	0.5
Slovakia	1.4	1.1	0.3	6.7	0.8	0.0
Greece	7.8	2.0	5.8	4.8	0.6	−0.8
Latvia	13.9	0.3	13.6	10.9	0.1	−8.5
Hungary	1.8	0.8	1.1	4.8	0.0	0.5
Poland	0.8	0.1	0.7	6.9	0.0	0.6
Croatia	1.7	0.4	1.2	13.6	0.2	0.7
Romania	0.6	0.3	0.3	18.4	0.2	0.0
Bulgaria	1.4	0.1	1.3	12.7	0.0	0.9

Source: Eurostat (2020c, 2020d, 2020e).
Note: Countries are listed in declining order of GDP per capita in 2019. The stock of foreign nationals living in a country (separately reported for EU-27 and non-EU-27 citizens) and the stock of a country's own citizens living elsewhere in the European Union are each reported as percentages of a country's current domestic population. For some countries, data on foreign citizens in the domestic population is unavailable for 2004, and data from the next available year is used instead.

The final two columns of Table 1 indicate the change in a country's immigrant share between 2004 and 2019, again differentiated by EU and non-EU citizens. It shows that immigrant shares increased in all but two countries, with several countries experiencing a growth of their foreign population share by 5 percentage points or more. The only exceptions are Latvia and Estonia, which saw many Russian nationals gain citizenship or returning to their home country. It is noteworthy that immigration from other EU countries was the main contributor to growing foreigner shares in most countries, especially those in Western Europe. Most of the European Union's Eastern member states only experienced modest increases in foreign population shares, which were often due to immigrants from outside the European Union, such as Ukrainians moving to Poland. Overall, the patterns of Table 1 clearly suggest intra-European labor flows from poorer to richer European countries, and especially from east to west.

We further investigate the regional patterns of immigrants' location choices in the Western European countries that are the main recipients of immigration in Europe. Figure 3 plots the nationality of the main foreigner group in a geographic region in the years 2018–19 based on data that we collected from individual countries. Our data set comprises 1,095 "NUTS-3" regions, as defined by the European Union.[6] There are 53 different nationalities that form the predominant group of foreigners in at least one of these 1,095 regions. Figure 3 does not try to display all of these nationalities separately, but indicates to which of seven different country groups the main foreign nationality belongs.

The blue areas of Figure 3 show that within many regions of Austria, Switzerland, France, Belgium and Luxembourg, the predominant foreigner group comes from an EU-15 or EFTA country, most often Portugal, Germany, Italy, or France. Out of the 15 countries that were part of the European Union prior to its eastern enlargement, *each* one accounts for the main group of foreigners in at least one region of another country.

If one looks at the blue, dark green, and light green areas, it is clear that in a majority of regions, the largest group of foreign nationals comes from another country within the European Economic Area. The dark green areas show that for a strikingly large number of regions, the main foreigner group hails from one of the countries that joined the European Union since 2004. Polish nationals form the largest immigrant group in most of the British Isles, and in parts of Scandinavia, Germany, and the Netherlands. In many regions of Italy and Spain, as well as in some areas of Austria and Southern Germany, the largest foreign groups instead are the citizens of Romania, which joined the European Union in 2007 and became the second most populous Eastern member state behind Poland.

---

[6]NUTS is an acronym for Nomenclature des Unités Territoriales Statistiques (Nomenclature of Territorial Units for Statistics), which is a hierarchical system of geographic regions that Eurostat uses for statistical purposes. The NUTS-3 units are defined as "small regions" that usually comprise populations between 150,000 and 800,000 individuals.

Figure 3

Origin Regions of Largest Foreign Nationality by NUTS-3 Geographic Region
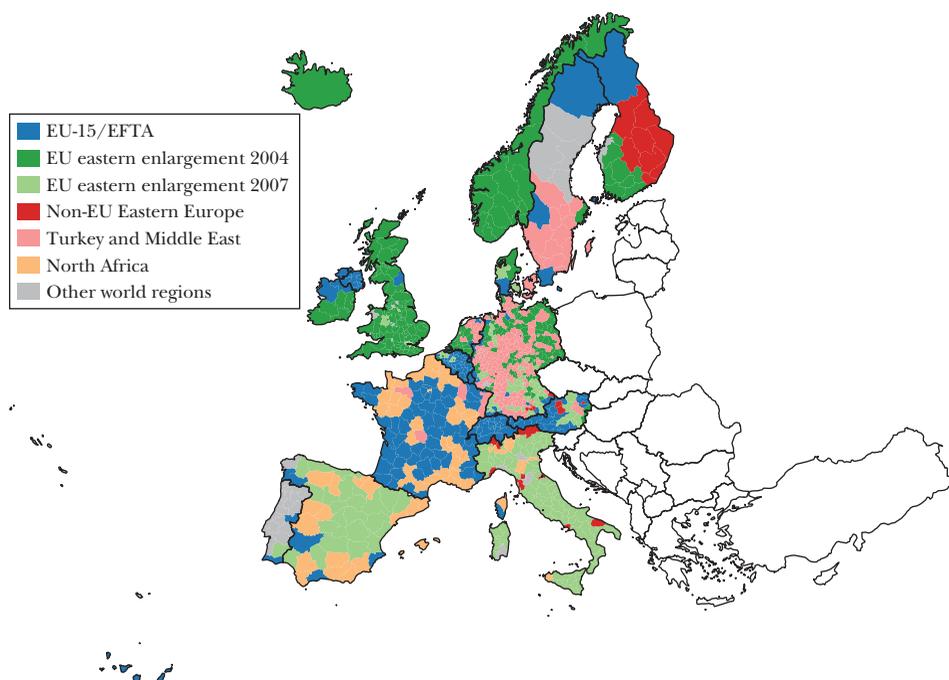
Source: All data is sourced from individual countries' statistical offices: Centraal Bureau vor de Statistiek (Netherlands) (2020a), Central Statistics Office (Ireland) (2017b), Danmarks Statistik (Denmark) (2020), Fundação Francisco Manuel dos Santos (Portugal) (2020), Hagstofa Islands (Iceland) (2020), Insituto Nacional de Estadística (Spain) (2020), Institut national de la statistique et des études économiques (France) (2020), Institut national de la statistique et des études économiques du Grand-Duché de Luxembourg (2020), Istituto Nazionale di Statistica (Italy) (2020), Office for National Statistics (United Kingdom) (2018a), Secrétariat d'etat aux migrations (Switzerland) (2020), Statbel (Belgium) (2020), Statistisches Bundesamt (Germany) (2020b), Statistik Austria (2020), Statistisk Sentralbyra (Norway) (2020), Statistiska Centralbyran (Sweden) (2020a), Tilastokeskus (Finland) (2020). © EuroGeographics for map with administrative boundaries.
Notes: The figure indicates the source region of the largest foreign nationality residing in each of 1,095 NUTS-3 regions of Western Europe, or in more aggregate NUTS-1 regions for Scotland and Northern Ireland. Population counts by nationality are measured on December 31, 2018, or January 1, 2019, if available, or at the latest available date otherwise.

Finally, the red, pink, orange, and gray areas on Figure 3 show regions where the main foreign nationality comes from a non-EEA country. In many regions of Germany, the Netherlands and southern Sweden, the dominant foreigner groups are Turkish or Syrian nationals, where the latter group includes many recently arrived refugees. Immigrants from North Africa, especially from Morocco and Algeria, form sizable communities in the Mediterranean countries Spain and France and to a lesser extent in Italy. Other source countries of immigrants that play a dominant role in a few regions include Brazilians in Portugal, Russians in Finland, Albanians in Italy, and Indians and Pakistanis in the United Kingdom.

What explains the location choices of different foreign nationalities that move to Western Europe? We investigated the choice of destination countries for the nine immigrant nationalities whose numbers in Western Europe grew the most between 2001 and 2018: Romania, Morocco, Syria, Poland, China, Bulgaria, Ukraine, Albania, and Russia. The nine panels of online Appendix Figure A2, available with this article at the *JEP* website, plot separately for each of these nationalities their initial percentage in the population of Western European countries in 2001, and the net inflow into these countries between 2001 and 2018.

For most immigrant nationalities, the initial stock and subsequent inflow to a destination country are positively correlated, which implies that immigrants tend to locate in countries that already host a sizable diaspora of the same nationality. This is the case, in particular, for non-European immigrants: Syrians moved primarily to Sweden and Germany, which already hosted relatively large proportions of Syrians in 2001, while Belgium remained a popular destination for immigrants from Morocco, and Italy for immigrants from China.[7]

Geographic distance also plays an apparent role in migrants' destination choices. For three of the nine main migrant nationalities, the net inflow from 2001 to 2018 was largest in the geographically closest Western European country, with Moroccans moving to Spain, Albanians moving to Italy, and Russians moving to Finland. Language distance arguably had a less important influence, because none of the nine sending countries shares a national language with a Western European country. However, Romanians did often move to Italy and Spain, whose languages are related to Romanian.

A particularly interesting pattern of migration is that for citizens of Poland, which is the largest Eastern European country that joined the European Union. In 2001, the share of Polish nationals was largest in Germany and Austria, the two Western European countries that are geographically closest to Poland. When Poland and other eastern European countries joined the European Union in 2004, Germany and Austria imposed transitional arrangements that deferred the opening of their labor markets to the new Eastern EU members to 2011. The only countries that immediately opened their labor markets to Eastern Europeans in 2004 were the United Kingdom, Ireland, and Sweden, while several other countries including Norway opened their markets in 2006. As a consequence of this staggered access to Western European labor markets, the largest net inflows of Polish immigrants relative to domestic population occurred in three countries that hosted few Polish nationals in 2001 but opened their markets early: Ireland, Norway, and the United Kingdom. Perhaps guided by that experience, these three countries no longer immediately allowed unrestricted immigration when Romania and Bulgaria joined

---

[7]An extreme counterexample to this pattern is the location choice of Ukrainians, whose net inflow was largest in Portugal, which was the country with lowest population share of that nationality in 2001. The number of Ukrainian citizens registered in Portugal grew from 71 individuals in 1996 to 62,448 individuals in 2002 (Fonseca and Pereira 2016). Most of these migrants benefited from a 2001 immigration law, which allowed individuals who had arrived with a tourist visa to gain a work permit after presenting an employment contract to authorities.

Figure 4
Share of Tertiary-Educated Individuals among Foreign Residents in 2000 and 2015
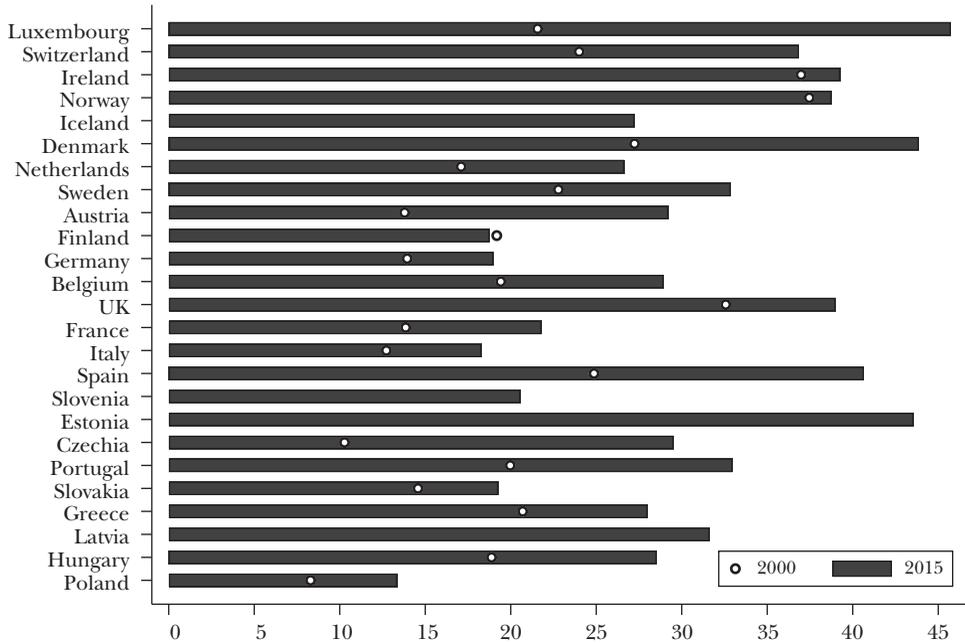
Note: Countries are listed in declining order of GDP per capita in 2015. All data refers to citizens of European OECD member countries who live in another European OECD member country and for whom education is known. Estonia, Iceland, Latvia, and Slovenia are included only in 2015 but not in 2000. Due to data availability, the initial share of foreigners with tertiary education is measured in 2005 instead of 2000 for Germany.

the European Union in 2007, but instead opened their markets to Romanians and Bulgarians only five to seven years later.

Migration by Educational Attainment

Much of the earlier migration from poorer to richer European countries, such as the flow of southern European guest workers to Germany in the 1960s and 1970s, involved unskilled workers who provided cheap labor in construction, factory jobs, or low-paid service occupations. However, globalization and technical change have raised the relative demand for high-skilled workers, particularly in countries with a comparative advantage in skill-intensive goods. As a consequence, worldwide migration to high-income countries has become more skill-biased in recent decades (Kerr et al. 2016).

Figure 4 shows the share of individuals with tertiary education in a country's foreign resident population both for the years 2000 and 2015. During this period, average education levels of immigrants increased in all countries but Finland, and

the increases were often large. Denmark, Luxembourg, Spain, and Switzerland all experienced a growth of the high-skill share among their foreign population by more than 12 percentage points. While the trends toward more highly educated immigrant populations is pervasive across countries, there remains large variation in the education levels of immigrants in different countries. For example, Spain (41 percent tertiary education share among immigrants) and the United Kingdom (39 percent) have relatively highly educated populations of foreigners, while Germany (19 percent), France (22 percent), and Italy (22 percent) have more low-skilled foreigner populations.

Although immigrant education levels have increased, immigrants remain less educated than natives in most European countries. In 2019, the tertiary education share in the EU-27 countries was 30 percent for foreign-born immigrants, but 35 percent for natives (Eurostat 2020g). Conversely, the share of individuals with at most a lower secondary education was considerably larger among the migrants (33 percent) than among the natives (17 percent).

A further differentiation of immigrants by source countries indicates that migrants within the European Economic Area possess slightly lower average education levels than natives, but higher education levels than immigrants from outside the EEA (Eurostat 2020g). Drawing on data from the 2007–2009 European Labor Force Survey, Dustmann and Frattini (2011) further report that individuals who moved between western EU countries had higher average educational attainment than the natives, while migrants who moved from the eastern to the western EU countries had lower education levels.

The data of Figure 4, which lists countries in declining order of their GDP per capita, suggest a weak positive correlation between a country's high-skill immigrant share and its income level. Moreover, countries that had higher income levels in 2000 also experienced a slightly larger growth in the high-skill immigrant share from 2000 to 2015. We thus find that migration not only flows from poorer to richer countries, but richer countries also tend to attract more skilled immigrants.

Equilibration of Labor Market Outcomes

The common European labor market can contribute to an equilibration of labor market outcomes across European countries. In theory, a complete removal of all mobility barriers should lead to factor price equalization. When production factors can be relocated without costs, the operation of market forces will attract workers to locations paying high wages and will induce firms to invest in locations where labor costs are low. In practice, however, markets are far from perfect. A broad set of mobility costs and frictions create substantial inertia. We discuss further below that even with open borders between European countries, obstacles to migration continue to exist due to different languages, heterogeneity in education, training and social security systems, as well as anti-immigrant attitudes of the native population and discrimination against immigrants. Given the presence of mobility

frictions, differences across countries in such dimensions as the skill composition of the labor force, industry composition, infrastructure, or institutional environment will continue to determine cross-country wage differences while making wage- and income-convergence a slow and long-lasting process; moreover, permanent differences in amenities offered by countries to workers and firms may inhibit full wage convergence.

Nonetheless, there is little doubt that the European integration process has substantially reduced mobility frictions, notably by giving foreign citizens within the European Economic Area the same legal access to a country's labor market that domestic citizens have. Head and Mayer (in this symposium) estimate that mobility costs within Europe fell rapidly in the 1960s, while reductions in these costs were more modest during the past two decades. Indeed, more than one-half of the EEA's current population live in the six founding members of the EEC for whom border-free mobility already became possible in the 1960s, and more than three-quarters live in countries that were part of the common labor market by the mid-1990s. Much of the removal of mobility barriers in Europe thus already occurred several decades ago.

Recent Wage Convergence in the European Labor Market

We discuss below the wage convergence across European countries between 2008 and 2018. Arguably the most important removal of mobility frictions during this period concerned the opening of the labor markets of the wealthier Western European countries to the citizens of the poorer Eastern European countries that have joined the European Economic Area since 2004. Therefore, one would, in particular, expect to see wage convergence between the Eastern and Western European countries.

Of course, forces other than migration will also affect convergence in European wage and income levels. There was arguably a large potential for catch-up growth in the Eastern European countries following their transition from communism to capitalism 30 years ago that would have led to some convergence even absent the common labor market. Moreover, by joining the European Union, the Eastern member states also gained access to the free movement of goods, capital, and services, and support through the European Union's spatial cohesion policy. That policy seeks to reduce economic disparities between countries and regions within the European Union. From 2014 to 2020, the European Union allocated about €645 billion, or one-third of its overall budget, to instruments such as subsidized infrastructure projects that support that cohesion policy (as discussed in this journal by Von Ehrlich and Overman 2020).

Yet despite these forces in favor of convergence, economic differences between the EU member countries remain remarkably large. In 2019, average labor costs in the European Union ranged from €6/hour in Bulgaria to €45/hour in Denmark (Eurostat 2020h). Although Denmark has the highest and Bulgaria has the lowest price level in the European Union, real wages still differ by a factor of 2.8 between the two countries (Eurostat 2020i).

Figure 5
Convergence in Real Wages across Countries
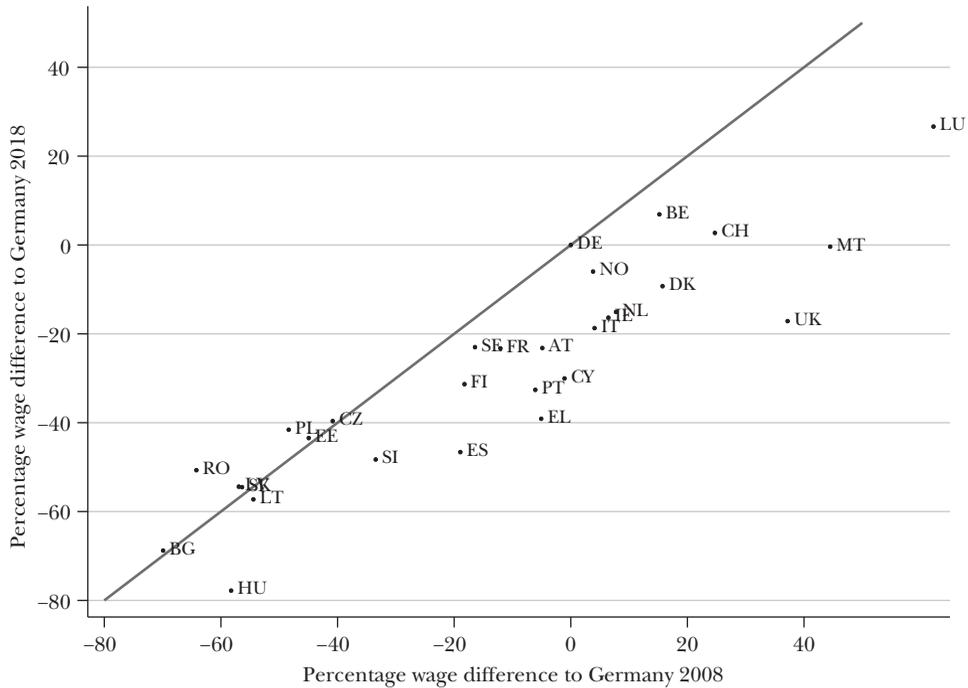
Here, we draw on microdata from EU-SILC (Statistics on Income and Living Conditions) to study convergence in real wages and in wages adjusted for skill levels. Our sample includes 253,894 workers in 2008 and 262,255 workers in 2018, who reside in 30 European countries. We regress, separately for each year, individuals' real log gross annual earnings on country fixed effects, and a set of control variables that includes a quartic in age and indicators for sex, marital status, and highest educational degree obtained. Germany is the reference country; hence, the coefficients for country fixed effects indicate countries' wage premia relative to Germany in the year under consideration.

Figure 5 plots coefficient estimates for the 2018 country fixed effects against those for the 2008 fixed effects, where we converted these effects from a log point to a percentage point scale. The graph also includes a 45-degree line. The evidence

shows that there was convergence in real wage levels. Consider the case of Romania ("RO") towards the bottom left corner of the figure. In 2008, the average Romanian worker earned 64 percent less than the average worker in Germany. However, that gap had shrunk to 51 percent ten years later. Indeed, for all Eastern European members of the EEA, except Hungary, Latvia, and Slovenia, the country indicators lie above the 45-degree line, as shown in the lower-left portion of the figure. These countries reduced their wage penalty relative to Germany. By contrast, the indicators for all Western European countries lie below the 45-degree line, as shown in the center and upper-right portion of the figure. Wage levels in these countries lost ground relative to Germany from 2008 to 2018.

A linear regression fit through the point cloud (not shown in the figure) would have a slope of 0.64. The slope of less than one implies that national wage levels had a greater dispersion in 2008 than in 2018: a wage difference which existed between any pair of countries in 2008 would be predicted to have shrunk by one-third by 2018.[8]

These wage estimates are broadly in line with recent evidence on convergence in GDP per capita across European countries and regions. Montfort (2020) finds convergence in per capita incomes between 2000 and 2008, which has slowed down substantially thereafter. Since 2008, overall convergence within the EU-28 has been weak, with countries of Eastern Europe slowly catching up, while there was some divergence within the EU-15. In sum, large income differences remain.

Static Earnings Gains from Migration

The persistent and large earnings differences across European countries suggest that migration within Europe is associated with high earnings gains for migrants. To shed light on the order of magnitude of these gains, we undertake a simple accounting exercise. First, we calculate the difference in earnings levels between origin country i and destination country j, $(w_j - w_i)$ for the year 2018, based on a regression of log real yearly earnings on worker characteristics and country fixed effects as in the analysis for Figure 5 above. Abstracting from the effects of immigrant selectivity (Borjas 1987) and immigrant assimilation (Chiswick 1978)—that is, the fact that immigrants typically face an earnings penalty initially and catch up only later on—differences in country fixed effects can serve as a measure for the earnings gain of a migrant moving from country i to country j.

---

[8] Details of the regression underlying Figure 5 and additional results are available in the online Appendix available with this article at the *JEP* website. In Appendix Figure A3, we repeat the same analysis based on an augmented cross-country wage regression that additionally controls for various characteristics of workers' jobs: weekly work hours, detailed occupation, and industry of employment. That setup seeks to isolate differences in countries' wage levels that cannot readily be explained by international differences in job types. Cross-country convergence is slightly weaker when we add these controls: a regression line through the point cloud of country fixed effect estimates has a slope of 0.68 rather than 0.64. Note that the estimated slope will be biased towards zero if countries' wage differences relative to Germany are measured with error. Therefore, one might interpret the slope estimate as lower bound for its actual value.

Second, we calculate the earnings gains from immigration for each origin-destination pair (i, j) as the product $M(i, j) \cdot (w_j - w_i)$, where $M(i, j)$ is the stock of migrants in destination j originating from country i.[9] These earnings gains are static in the sense that they take wage levels as given, and abstract from any impact of migration on wages in the origin or destination countries.

Based on this calculation, we estimate that 12.7 million intra-EU migrant workers obtain an average earnings gain of about €6,500 per year each (all numbers adjusted by purchasing power parity throughout). The product of these numbers yields an aggregate static earnings gain from within-EU migration of €83.2 billion, or 0.5 percent of EU-wide GDP. The bulk of this benefit, €67.9 billion, accrues to migrants from Eastern Europe, whose earnings gains amount to 2.8 percent of eastern EU countries' GDP. For Bulgaria, which is the poorest member country of the European Union, the static migration gain is largest at 8.0 percent of GDP.

This basic calculation assumes that the gain from migrating from country to country is the same for all workers of a given broad education group. In reality, the potential gains from migration may, however, vary across workers, and it is plausible that those who stand to gain more will be more likely to migrate. By not taking into account this selection effect, one will tend to underestimate the gains from migration. In particular, while our simple calculation implies an earnings loss for every worker who moves from a richer to a poorer country, it is possible that at least some of these workers in reality earn more in the low-wage host-country than in their high-wage home-country. If we only take into account migration flows from poorer to richer countries, the EU-wide gains from migration are indeed larger, amounting to €97.4 billion or 0.6 percent of EU GDP. For eastern EU countries, the gains from migration are, unsurprisingly, barely affected, because almost all migrants from Eastern Europe move to a richer EU country.

Another potential source of bias in our baseline calculation stems from the fact that foreign citizens often obtain lower wages in a destination country than domestic citizens. By ignoring that pattern, one will tend to overestimate the earnings gains from migration. Indeed, when we account for such wage penalties by calculating separate wage levels in a country for domestic and foreign citizens, then the gains from migration are substantially smaller. They amount to 0.2 percent of EU GDP for EU-wide migration, to 1.7 percent of GDP for the member states in Eastern Europe, and to 5.7 percent of GDP for the poorest country, Bulgaria.

Clemens (2011) in this journal reviews a broader literature on the potential gains from reducing worldwide barriers to labor mobility. While a complete removal of such barriers could generate gains of more than 100 percent of worldwide GDP

---

[9] We calculate country-specific wage premia and earnings gains separately for migrants with and without tertiary education. The gains are adjusted by purchasing power parity exchange rates. Online Appendix Figure A4 shows that in many countries, wage differences relative to Germany are larger for highly educated workers than for less educated ones. Our calculation also takes into account that earnings gains from migration accrue not to all migrants, but only to working-age individuals who are employed. The online Appendix provides further detail on this computation of gains from migration.

according to some estimates, the realization of such gains would require that more than half of the world population moves to another country. With partial reductions of mobility barriers that lead to a migration of about 1–2 percent of the world population, world GDP could still grow by about 1–2 percent. In comparing such calculations to migration gains for Europe, it is important to note that income differentials within Europe are much smaller than worldwide, which leads to smaller potential gains from migration.

Indeed, our simple quantification of migration gains can be used to highlight how gains from migration depend on both migration rates and earnings differences between countries. Consider first the case of Bulgaria. The number of Bulgarian workers in other EU countries corresponds to about 8 percent of Bulgaria's population, and the average migration gain per Bulgarian worker is about equal to the country's per-capita GDP. As a consequence, we obtain a migration gain of 8 percent of GDP for Bulgaria in our baseline calculation. When we look instead at the entire European Union, both the fraction of migrant workers (about 2.5 percent) and the average gain per worker (about 20 percent of per-capita EU GDP) are substantially lower than in the Bulgarian case, and in combination result in the much smaller migration benefit of 0.5 percent of EU GDP.

Earnings Effects of Immigration on Host-Country Wages

The simple accounting exercise above calculated earnings gains from within-EU migration based on the assumption that wages in the involved countries are not themselves affected by immigration. This is a strong assumption, but perhaps somewhat less unreasonable given the large and highly persistent cross-country variation in real wages across European countries.

One possible explanation for that persistence in wage differentials is that the labor flows within the common European labor market are not large enough to create a stronger convergence in wage levels. Another possible explanation is that labor markets adjust to immigration primarily through an adjustment of employment, rather than an adjustment of wages. For example, Glitz (2012) looks at the large immigration flow into Germany of 2.8 million ethnic Germans from Eastern Europe and the former Soviet Union during the 1990s and early 2000s. He finds no effect on wages, but a large employment effect: for every ten immigrants who find a job, three native workers become unemployed. A related study by Dustmann, Schönberg, and Stuhler (2017) analyzes a local labor supply shock in a German border region when workers from nearby Czechia were allowed to enter the country. It finds a moderate decline in the German wage but a large negative response in local native employment.

The intuitive conjecture that migration should equilibrate wages and employment rates rests on the implicit assumptions that labor is homogeneous and that labor demand is constant. But if labor is heterogeneous and there is little substitutability between immigrant and native workers, then a migrant inflow will generate little downward pressure for the wages of natives. Conversely, immigrants may contribute to firm growth by filling important labor shortages or by contributing

to innovation, in which case, immigration may trigger an increase in labor demand that raises the native wage level.

With these ideas in mind, certain areas of Switzerland offer an interesting case study for the effects of European labor market integration. The Swiss had rejected membership in the European Economic Area in a 1992 referendum, and only became part of the common European labor market in 2005 after a set of bilateral agreements with the European Union. From 2000 to 2019, immigration increased the share of foreigners in the Swiss workforce by more than 8 percentage points. This surge in immigrant workers included many workers who reside in neighboring regions of Italy, France, and Germany, and who commute daily to Switzerland in order to take advantage of the elevated Swiss wages. The number of workers employed in Switzerland but residing in a neighboring country almost tripled since 2000 and now accounts for an astonishing 6.5 percent of the Swiss labor force.

Cross-border work in Switzerland is particularly important in the cantons of Geneva near the French border and Ticino near the Italian border, where cross-border workers account for 26 and 29 percent of all workers in those cantons, respectively. In these cases, frictions to cross-border labor mobility seem very limited. In particular, there are no restrictions arising from language differences (Geneva is a French-speaking canton; Ticino is an Italian-speaking canton), and cross-border transportation systems are well developed. Several recent studies explore how the increase in cross-border work affected the local labor markets of both Switzerland, where labor supply increased dramatically, as well as the border regions of France and Italy, which lost many workers to Switzerland.

Beerli et al. (2021) find that the increase in cross-border workers in the most strongly exposed border regions of Switzerland left wages and employment of native Swiss workers largely unchanged. Indeed, wages of university-educated natives even increased. It appears that migration allowed highly productive and skill-intensive firms to close their labor shortages. Conversely, the French and Italian border regions lost a sizeable fraction of their employees to Swiss firms. For the French border regions, Hafner (2021) finds that the wages of low-skilled workers were slightly rising, while wages of high-skilled workers remained unaffected. Dicarlo (2020) shows that Italian firms in the border region faced substantial labor shortages after large numbers of Italian workers took up jobs in nearby Switzerland. In particular, Italian firms in high-skill sectors in the border region struggled to compensate for this loss in labor supply. Nevertheless, wages in these firms declined, most likely because the most productive workers went to Switzerland.

Taken together, these papers suggest that labor market integration between Switzerland and its neighbors did not decrease—and perhaps even increased—the wage differences across national borders. Various studies have also found positive wage effects of immigrants on natives in other European countries. For instance, Dustmann, Frattini, and Preston (2013) find that, on average, immigration in the UK slightly increased the average wage of native workers, though wages responded

differentially along the wage distribution (some wage declines below the twentieth percentile of the wage distribution but modest wage gains in the upper ranks of the distribution). Ortega and Verdugo (2014) show that immigration into France raised the wage of French workers by fostering a reallocation of the native workers to better-paying occupations.

The general message from all these studies is that migration flows may have surprisingly weak effects on wages. Despite increasing migration flows within Europe, an equilibration of wage levels across countries does not seem near.

Obstacles to Migration and European Labor Market Integration

The labor market of the European Economic Area remains considerably less integrated than the US labor market and has much lower migration rates. A proximate reason for these relatively modest migration rates in Europe is that labor market outcomes for migrants are often worse than those of similarly educated natives. Some citizens of Europe's poorer countries would likely struggle to obtain adequate jobs if they moved to a richer country, and their financial gain from moving would thus be considerably smaller than suggested by the large international wage differences indicated in Figure 5. Algan et al. (2011) review the labor market performance of immigrants in Europe's three largest economies—Germany, France, and the United Kingdom—and conclude that immigrants do worse than natives in terms of employment rates and earnings, after controlling for education, potential experience and regional location. The immigrant-native gaps appear quite persistent across first- and second-generation immigrants (that is, native-born children of foreign-born parents).

Importantly, immigrants' labor market performance varies widely across immigrant groups. While migrants from other Western European countries have fairly similar outcomes than natives, very large gaps exist for immigrants from outside Europe of different races and ethnicities, such as Africans in France, or Bangladeshis and Pakistanis in the United Kingdom. Eastern Europeans, and in some cases southern Europeans such as Greeks or Italians in Germany, also do worse than the natives. Calmfors and Sánchez Gassen (2019) show that immigrants' employment prospects are substantially below those of natives even in the egalitarian Nordic countries.

Language and Culture

Europe's remarkably large heterogeneity in languages is one reason why immigrants may struggle to gain a foothold in another country's labor market. The European Union alone lists 24 different official languages, and the non-EU members of the common labor market add another three. A lack of proficiency in the destination country's language not only limits immigrants' ability to find jobs quickly but can also reduce productivity in the workplace and social inclusion. A large literature has documented that poor language proficiency has a sizable

negative effect on labor earnings of immigrants (Chiswick and Miller 2014). Other research suggests that language differences between the origin and destination countries constitute a barrier for migration. Adserà and Pytliková (2015) show that in a panel of OECD countries, migration flows are stronger between countries that share the same language. Moreover, English-speaking countries generally receive greater migrant inflows, which is likely due to the widespread teaching of English as a foreign language. In the European Union, 96 percent of all students in upper secondary education learn English as a foreign language, while the fractions of students learning Spanish, French, and German are just 26 percent, 22 percent, and 20 percent, respectively (Eurostat 2020g).

Language can also more broadly proxy for local culture, and migrants across language borders may have to learn not only a new language but also to familiarize themselves with local practices of interpersonal interaction and labor market behavior. Consistent with such an interpretation of language as a proxy for culture, Eugster et al. (2017) show that workers' job search behavior differs notably across nearby German-speaking and French-speaking regions in Switzerland that share the same formal labor market institutions.

Education, Training, and Social Security

Certain institutional features may also hinder the smooth integration of immigrants into host country labor markets. European education and occupational training systems are organized and administered at the national level. Because these systems differ across countries, skilled immigrants often face limitations to enter the occupation in which they were trained at home. In some cases, employers may have difficulty assessing educational credentials that were acquired abroad; in others, occupational licensing rules make it difficult to get formal recognition of occupational certificates acquired abroad.

Tertiary education is one area where standards have been harmonized. The 1999 Bologna declaration was signed by 29 European countries (the EU-28 except Cyprus, plus Norway and Switzerland). In follow-up agreements, the "Bologna process" was opened to other countries, including those of the former Soviet Union, former Yugoslavia, and Turkey, and now includes 48 countries that form the European Higher Education Area. In this agreement, countries coordinated on adopting a system of comparable degrees, similar study cycles (undergraduate/graduate), and a system of portable study credits. Furthermore, there is an agreement to promote international mobility of students and teaching staff and to harmonize the standards and quality of study programs (Huisman et al. 2012). By 2018, 1.3 million students enrolled in tertiary programs across the EU-27 came from abroad, with 44 percent coming from other European countries. Germany, France, Italy, and the Netherlands attracted more than half of these foreign students (Eurostat 2020g).

EU legislation has also sought to standardize and facilitate the process of occupational recognition, yet significant barriers remain. Koumenta et al. (2014) document that access to more than 800 occupations is regulated in at least one EU member state, with these occupations covering up to 24 percent of the EU

labor force. They show that intra-EU migrants are less likely than natives to enter a profession subject to licensing. Further analyses show that occupational recognition has a significant effect on wages. Brücker et al. (2020), studying the impact of occupational recognition in Germany, find that three years after obtaining recognition of their occupational credentials, immigrants earn 20 percent higher wages and are 25 percent more likely to be employed than similar immigrants who never applied for recognition.[10] Obstacles to occupational recognition likely contribute to occupational downgrading, where immigrants work in jobs that are inferior to their previous education and labor market experience (Dustmann, Frattini, and Preston 2013).

An additional mobility barrier concerns the large heterogeneity in social insurance rights across European countries. These rights—including old-age pensions, unemployment payments, and government-financed healthcare services—are determined at the national level, and programs differ strongly across countries. For migrants, it is not always obvious whether rights acquired in one country are transferable to another country. For instance, a worker who moves frequently across countries and works for only short periods in each of them may not satisfy any country's minimum qualifying period that is required to gain access to an old-age pension. "Coordination Regulations" have been established to facilitate the portability of social insurance rights across countries and to prohibit discrimination against immigrants or against return migrants who have since left a country (European Commission 2019).

Discrimination and Anti-immigrant Attitudes

Another explanation for immigrants' relative lack of labor market success is discrimination in the labor market. There is ample evidence from Europe and elsewhere for discrimination against racial and ethnic minorities in the labor market, which is reviewed in recent surveys by Bertrand and Duflo (2017) and Neumark (2018).

In the context of migration within Europe, differences in national origins are not necessarily visible from workers' physiques (used in audit studies of in-person job applicants) or from workers' names (in correspondence studies based on submissions of written job applications). Thus, one recent study that explicitly investigates discrimination by nationality uses data from an online platform of the Swiss public employment service that connects job seekers with recruiters (Hangartner, Kopp, and Siegenthaler 2021). On this platform, recruiters observe not only the names but also the nationalities and language skills of job seekers. Holding constant other observables, job seekers of non-European origin are 13 to 19 percent less likely to be contacted by recruiters than Swiss nationals. For migrants within the common European labor market, penalties are smaller and range from zero for southern Europeans (which include Italians who form the

---

[10] In a US context, Kleiner and Krueger (2013) estimate that 29 percent of jobs are subject to occupation licensing rules and that licensing is associated with 18 percent higher wages.

largest group of foreigners in Switzerland) to 6 percent, both for immigrants from the northwest and east of the continent. Most of these penalties disappear when immigrants are naturalized, although recruiters may still infer the foreign roots of applicants based on their names and language skills in some cases (Kopp, Siegenthaler, and Hangartner 2020). The nationality of job applicants thus appears to play an important role in labor market discrimination, rather than just the ethnicity. Åslund, Hensvik, and Nordström Skans (2014) additionally show that hiring chances of immigrants in Sweden are significantly lower in firms whose managers are born in Sweden instead of abroad, which suggests that discrimination may result from homophily.

The free migration of labor within Europe is arguably the most politically controversial element of the common European market. Alfano et al. (2016) argue that the United Kingdom's lack of control over immigration from the European Economic Area became the single most important argument in favor of the "Brexit" referendum, which eventually led to the United Kingdom's exit from the common market. However, support for Brexit was highest not in those regions that had received the most immigration in previous years, but in regions that experienced economic decline due to rising international trade competition (Colantone and Stanig 2018).

While it is unclear whether immigration has adverse impacts on the labor market outcomes of natives (Dustmann, Schönberg, and Stuhler 2016), migrants affect natives also by changing the composition of nationalities, languages, and cultures in neighborhoods, workplaces and schools. Card, Dustmann, and Preston (2012) find that concerns related to such compositional amenities are 2–5 times more important than concerns about the labor market in order to explain people's attitudes towards migrants.

Despite the United Kingdom's exit from the common market and the rise of anti-immigrant sentiment in some European countries, attitudes of the general public towards immigration have not become more skeptical during the last two decades in most countries. We compiled data from the 2004 and 2018 European Social Survey, which asked respondents "to what extent do you think your country should allow people of the same race or ethnic group as most of your country's people to come and live here?" The fraction of survey respondents who answered either "allow many" or "allow some" (instead of "allow few" or "allow none") increased in 13 out of 14 Western European countries, from an average of 66 to 77 percent, with declining support for immigration being observed only in Italy. In the six countries of Eastern Europe included in the surveys, support for immigration changed modestly from an average of 59 to 58 percent, with declines in Czechia, Poland, and Slovakia (for details of the survey results, see online Appendix Figure A5).

Inflexible Domestic Labor Markets

While obstacles to labor migration across European countries exist, it is worth pointing out that job-to-job mobility is also quite low *within* many European countries. The same reasons that prevent workers from changing jobs domestically

may also keep them from moving internationally. In particular, southern European countries tend to have strict employment protection regulations that require employers to pay sizable compensations to workers in case of layoffs. Such measures strongly reduce worker mobility across jobs (Martin and Scarpetta 2012). Alesina et al. (2015) also argue that the cultures of southern European countries value close family ties more strongly than cultures in northern European or Anglo-Saxon countries. In a culture with strong family ties, many adults do not want to move far away from their parents and relatives, which limits spatial mobility even if migration would be financially gainful.

Conclusions

We are still far from a common European labor market. In a 2014 survey conducted by the German think tank IZA, among 284 European labor economists, nearly three-quarters *disagreed* with the statement that "the single European labor market is largely achieved" (Krause-Pilatus, Rinne, and Zimmermann 2014). Despite the removal of legal barriers to labor mobility, large differences in labor market outcomes across European countries remain.

Of course, most domestic labor markets—including the US labor market—are segmented into geographic local labor markets where localized shocks can lead to fairly persistent differentials in wage and unemployment levels (Moretti 2011; Autor, Dorn, and Hanson 2021). However, migration rates within the European labor market are much smaller than in the United States, despite larger geographic differentials in labor market outcomes across European regions, and notwithstanding that Europe covers a larger population distributed over a much smaller land area. As noted, some of the remaining obstacles to a more integrated European labor market include heterogeneity of Europe in terms of languages and cultures; national regulations related to education, training, and employment conditions; and discrimination against migrants.

National borders are no longer legal barriers to labor migration, but they remain important for Europeans' self-identification. Four in seven EU citizens (57 percent) feel very attached to their own country, while only one in seven (14 percent) feel very attached to the European Union (European Commission 2018). The United Kingdom's departure from the common market—which was partly driven by concerns about migration—makes clear that further European labor market integration cannot be taken for granted. While there is currently no indication that other countries will soon follow the United Kingdom's path of leaving the European Union, it is also unlikely that the European labor market will substantially grow over the next decade through the accession of new member states. The European Union has opened membership negotiations with five countries, but the negotiations with the largest candidate country (Turkey) have now been frozen for many years, and the other four countries (Albania, Montenegro, North Macedonia, and Serbia) would add less than 3 percent to the population of the European Economic Area. A

further integration of the European labor market may thus more likely result from the European Union's efforts to harmonize or coordinate national regulations in order to reduce obstacles to migration, and from continued migration of workers from Eastern to Western Europe.

References

Adserà, Alícia, and Mariola Pytliková. 2015. "The Role of Language in Shaping International Migration." *Economic Journal* 125 (586): 49–81.

Alesina, Alberto, Yann Algan, Pierre Cahuc, and Paola Giuliano. 2015. "Family Values and the Regulation of Labor." *Journal of the European Economic Association* 13 (4): 599–630.

Alfano, Marco, Christian Dustmann and Tommaso Frattini. 2016. "Immigration and the UK: Reflections after Brexit." CReAM Discussion Paper 1623.

Algan, Yann, Christian Dustmann, Albrecht Glitz, Alan Manning. 2010. "The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom." *The Economic Journal* 120 (542): 4–30.

Åslund, Olof, Lena Hensvik, and Oskar Nordström Skans. 2014. "Seeking Similarity: How Immigrants and Natives Manage in the Labor Market." *Journal of Labor Economics* 32 (3): 40–41.

Autor, David, David Dorn, and Gordon Hanson. 2021. "On the Persistence of the China Shock." Unpublished.

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." A*merican Economic Review* 111(3): 976–1012..

Bertrand, Marianne, and Esther Duflo. 2016."Field Experiments on Discrimination." NBER Working Paper 22014.

Bolt, Jutta, and Jan Luiten van Zanden. 2020. "Maddison Project Database, version 2020." https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2020 (accessed November 15, 2020).

Borjas, George J. 1987. "Self-Selection and the Earnings of Immigrants." *American Economic Review* 77 (4): 531–53.

Brücker Herbert, Albrecht Glitz, Adrian Lerche, and Agnese Romiti. 2018. "Occupational Recognition and Immigrant Labor Market Outcomes." IZA Discussion Paper 12030.

Bureau of Labor Statistics. 2019. *Regional and State Unemployment–2019 Annual Averages.* US Bureau of Labor Statistics.

Calmfors, Lars, and Nora Sánchez. 2019. "Integrating Immigrants into the Nordic Labour Markets: Background, Summary and Policy Conclusions." In *Integrating Immigrants into the Nordic Labour Markets*, edited by Nordic Council of Ministers, 9–36. Denmark: Nordic Council of Ministers.

Card, David, Christian Dustmann, and Ian Preston. 2012. "Immigration, Wages and Compositional Amenities." *Journal of the European Economic Association* 10 (1): 78–119.

Centraal Bureau vor de Statistiek. 2020. "Bevolking; geslacht, leeftijd, nationaliteit en regio, 1 januari." StatLine (Netherlands). https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84727NED/table?ts=1616414082872 (accessed November 15, 2020).

Central Statistics Office. 2017a. "E7002: Population Usually Resident and Present in the State 2011 to

2016 by Nationality, Sex, County of Usual Residence and Census Year." PxStat Database (Ireland). https://data.cso.ie/ (accessed November 15, 2020).

Central Statistics Office. 2017b. "B0439 - 2002 Population Usually Resident and Present in their Usual Residence." PxStat Database (Ireland). https://data.cso.ie/# (accessed February 18, 2021).

Chiswick, Barry R. 1978. "The Effect of Americanization on the Earnings of Foreign-born Men." *Journal of Political Economy* 86 (5): 897–921.

Chiswick, Barry R., and Paul W. Miller. 2014. "International Migration and the Economics of Language." In *Handbook of the Economics of International Migration*, Vol. 1A, edited by Barry R. Chiswick and Paul W. Miller, 211–69. Amsterdam: Elsevier.

Clemens, Michael A. 2011. "Economics and Emigration: Trillion-Dollar Bills on the Sidewalk." *Journal of Economic Perspectives* 25 (3): 83–106.

Colantone, Italo, and Piero Stanig. 2018. "Global Competition and Brexit." *American Political Science Review* 112 (2): 201–18.

Danmarks Statistik. 2020. "Population at the First Day of the Quarter by Region, Sex, Age and Group of Citizenship (2008Q1_2020Q4)." StatBank Denmark. https://www.statbank.dk/statbank55/SelectVarVal/Define.asp?MainTable=FOLK1D&PLanguage=1 (accessed November 15, 2020).

De la Rica, Sara, Albrecht Glitz, and Francesc Ortega. 2015. "Immigration in Europe: Trends, Policies, and Empirical Evidence." In *Handbook of the Economics of International Migration*, Vol. 1B, edited by Barry R. Chiswick and Paul W. Miller, 1303–62. Amsterdam: Elsevier.

DiCarlo, Emanuele. 2020. "How Do Firms Adjust to Negative Labor Supply Shocks? Evidence from Migration Outflows." Unpublished.

Dustmann, Christian, and Tommaso Frattini. 2011. "Immigration: The European Experience", IZA Discussion Paper 6261.

Dustmann, Christian, Tommaso Frattini, and Ian Preston. 2013. "The Effect of Immigration along the Wage Distribution." *Review of Economic Studies* 80 (1): 145–73.

Dustmann, Christian, and Ian P. Preston. 2019. "Free Movement, Open Borders, and the Global Gains from Labor Mobility." *Annual Review of Economics* 11: 783–808.

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.

Dustmann, Christian, Uta Schönberg, and Jan Stuhler. 2017. "Labor Supply Shocks, Native Wages, and the Adjustment of Local Employment." *Quarterly Journal of Economics* 132 (1): 435–83.

Eugster, Beatrix, Rafael Lalive, Andreas Steinhauer, and Josef Zweimüller. 2017. "Culture, Work Attitudes and Job Search: Evidence from the Swiss Language Border." *Journal of the European Economic Association* 15 (5): 1056–1100.

European Commission. 2018. "European Citizenship." *Standard Eurobarometer* 89 (Spring 2018): 36–39.

European Commission. 2019. *Coordination of Social Security Systems at a Glance, 2019 Statistical Report.* Brussels: European Commission.

Eurostat. 2020a. "Unemployment by Sex and Age—Annual Data Database." Eurostat. https://ec.europa.eu/eurostat/web/products-datasets/-/UNE_RT_A (accessed November 15, 2020).

Eurostat. 2020b. "Migration and Migrant Population Statistics." https://ec.europa.eu/eurostat/statistics-explained/index.php/Migration_and_migrant_population_statistics.

Eurostat. 2020c. "Population on 1 January by Age and Sex Database." Eurostat. https://ec.europa.eu/eurostat/web/products-datasets/-/DEMO_PJAN (accessed November 15, 2020).

Eurostat. 2020d. "Population on 1 January by Age Group, Sex and Citizenship Database." Eurostat. https://ec.europa.eu/eurostat/web/products-datasets/-/MIGR_POP1CTZ (accessed November 15, 2020).

Eurostat. 2020e. "EU and EFTA Citizens Who Are Usual Residents in Another EU/EFTA Country as of 1 January Database." Eurostat. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=migr_pop9ctz&lang=en (accessed November 15, 2020).

Eurostat. 2020f. "Purchasing Power Adjusted GDP per Capita Database." Eurostat. https://ec.europa.eu/eurostat/web/products-datasets/-/sdg_10_10 (accessed November 15, 2020).

Eurostat. 2020g. "Learning Mobility Statistics." https://ec.europa.eu/eurostat/statistics-explained/index.php/Learning_mobility_statistics.

Eurostat. 2020h. "Hourly Labour Costs." https://ec.europa.eu/eurostat/statistics-explained/index.php/Hourly_labour_costs.

Eurostat. 2020i. "Comparative Price Levels of Consumer Goods and Services." https://ec.europa.eu/eurostat/statistics-explained/index.php/Comparative_price_levels_of_consumer_goods_and_services.

Eurostat. 2020j. "Purchasing Power Parities (PPPs), Price Level Indices and Real Expenditures for ESA 2010 aggregates Database." Eurostat. https://ec.europa.eu/eurostat/web/products-datasets/-/PRC_PPP_IND (accessed November 15, 2020).

Eurostat. 2020k. "European Union Statistics on Income and Living Conditions (EU-SILC) Database." Eurostat. Restricted access with data contract (accessed November 15, 2020).

Fihel, Agnieszka, Anna Janicka, Paweł Kaczmarczyk, and Joanna Nestorowicz. 2015. *Free Movement of Workers and Transitional Arrangements: Lessons from the 2004 and 2007 Enlargements.* Warsaw: Centre of Migration Research, University of Warsaw.

Fonseca, Maria L., and Sonia Pereira (2016). Migration of Ukrainian Nationals to Portugal: The Visibility of a New Migration Landscape. In: Fedyuk, Olena and Marta Kindler, eds.: "Ukrainian Migration to the European Union." Cham: Springer.

Fundação Francisco Manuel dos Santos. 2020. "Foreign Population with Legal Resident Status by Nationalities." PORDATA Base de Dados Portugal Contemporâneo. https://www.pordata.pt/en/Portugal/Foreign+population+with+legal+resident+status+total+and+by+certain+nationalities-24 (accessed November 15, 2020).

Glitz, Albrecht. 2012. "The Labor Market Impact of Immigration. A Quasi-Experiment Exploiting Immigrant Location Rules in Germany." *Journal of Labor Economics* 30 (1): 175–213.

Guiso, Luigi, Helios Herrera, Massimo Morelli and Tommaso Sonno. 2020. "Demand and Supply of Populism." Unpublished.

Hafner, Flavio. 2021. "Labor Market Competition, Wages and Worker Mobility." Working Paper, Pompeu Fabra University.

Hagstofa Islands. 2020. "Population by Municipality, Sex, Citizenship and Quarters 2010–2020." Statistical Database (Iceland). http://px.hagstofa.is/pxen/pxweb/en/Ibuar/Ibuar__mannfjoldi__1_yfirlit__arsfjordungstolur/MAN10001.px (accessed November 15, 2020).

Hangartner, Dominik, Daniel Kopp, and Michael Siegenthaler. 2021. "Monitoring Hiring Discrimination through Online Recruitment Platforms." Nature: 589 (7843): 572–76.

Huisman, Jeroen, Clifford Adelman, Chuo-Chun Hsieh, Farshid Shams and Stephen Wilkins (2012). Europe's Bologna process and its impact on global higher education. In D.K. Deardorff, H. de Wit, J.D. Heyl, and T. Adams (eds.), *The SAGE Handbook of International Higher Education*, pp. 81-100. Thousand Oaks: Sage Publications.

Instituto Nacional de Estadística. 2020. "Población extranjera por Nacionalidad, provincias, Sexo y Año." Instituto Nacional de Estadística (Spain). https://www.ine.es/jaxi/Tabla.htm?path=/t20/e245/p08/l0/&file=03005.px (accessed November 15, 2020).

Institut national de la statistique et des études économiques. 2020. "Etrangers - Immigrés en 2017." Recensement de la population (France). https://www.insee.fr/fr/statistiques/zones/4515432 (accessed November 15, 2020).

Institut national de la statistique et des études économiques du Grand-Duché de Luxembourg. 2020. "Etat de la population." Le portail des statistiques. https://statistiques.public.lu/stat/ReportFolders/ReportFolder.aspx?IF_Language=fra&MainTheme=2&FldrName=1 (accessed November 15, 2020).

International Migration Institute. 2015. "DEMIG C2C." Demig data. https://www.migrationinstitute.org/data/demig-data/demig-c2c-data (accessed September 6, 2020).

Istituto Nazionale di Statistica. 2020. "Resident Foreigners on 1st January—Citizenship: Italy, Regions, Provinces—Country of Citizenship." Statbase. http://dati.istat.it/viewhtml.aspx?il=blank&vh=0000&vf=0&vcq=1100&graph=0&view-metadata=1&lang=it&QueryId=19675&metadata=DCIS_POPSTRCIT1 (accessed November 15, 2020).

Jeannet, Anne Marie. 2017. "Political Distrust in Europe: the Impact of Immigration and the Global Economic Crisis." DONDENA Working Paper 102.

Kerr, Sari Pekkala, William Kerr, Çağlar Özden, and Christopher Parsons. 2016. "Global Talent Flows." *Journal of Economic Perspectives* 30 (4): 83–106.

Kleiner, Morris M., and Alan B. Krueger. 2013. Analyzing the extent and influence of occupational licensing on the labor market. *Journal of Labor Economics* 31, no. 2:S173– S202.

Kopp, Daniel, Michael Siegenthaler, and Dominik Hangartner. 2020. "Host Country Citizenship Reduces Hiring Discrimination against Immigrant Minorities." Unpublished.

Koumenta, Maria, Amy Humphris, Morris Kleiner, and Mario Pagliero. 2014. *Occupational Regulation in the EU and UK: Prevalence and Labour Market Impacts.* London: Queen Mary University of London.

Krause-Pilatus, Annabelle, Ulf Rinne, and Klaus F. Zimmermann. 2014. "How Far Away Is a Single

European Labor Market?" IZA Discussion Paper 8383.

Margalit, Yotam. 2019. "Economic Insecurity and the Causes of Populism, Revisited." *Journal of Economic Perspectives* 33 (4): 152–70.

Martin, John P., and Stefano Scarpetta. 2012. "Setting It Right: Employment Protection, Labour Reallocation and Productivity." *De Economist,* 160: 89–116.

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

Montfort, Philippe. 2020. "Convergence of EU Regions Redux: Recent Trends in Regional Disparities." Directorate-General for Regional and Urban Policy Working Paper 02/2020.

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

Neumark, David. 2018. "Experimental Research on Labor Market Discrimination." *Journal of Economic Literature* 56 (3): 799–866.

OECD. 2020. "Database on Immigrants in OECD and non-OECD Countries: DIOC." OECD. http://www.oecd.org/els/mig/dioc.htm (accessed November 15, 2020).

Office for National Statistics. 2018a. "Table 2.1: Population in the UK, Excluding Some Residents in Communal Establishments, by Nationality; UK by Countries and, within England, London Boroughs, Metropolitan Districts, Unitary Authorities and Non-Metropolitan Districts; and Unitary Authorities within Wales." National Statistics (United Kingdom). https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/datasets/populationoftheunitedkingdombycountryofbirthandnationalityunderlyingdatasheets (accessed November 15, 2020).

Office for National Statistics. 2018b. "Table 2.3: Non-British Population in the UK, Excluding Some Residents in Communal Establishments, by Sex, by Nationality." National Statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/datasets/populationoftheunitedkingdombycountryofbirthandnationalityunderlyingdatasheets (accessed November 15, 2020).

Ortega, Javier, and Gregory Verdugo. 2014. "The Impact of Immigration on the French Labor Market: Why so Different?" *Labour Economics* 29: 14-27.

Secrétariat d'Etat aux migrations SEM. 2020. "Statistique sur les étrangers, janvier 2019." Swiss Federal Department of Justice and Police. https://www.sem.admin.ch/sem/fr/home/publiservice/statistik/auslaenderstatistik/archiv/2019/01.html (accessed November 15, 2020).

Statbel. 2020. "Population par commune selon la nationalité et le sexe depuis 1992." Service Public Fédéral Belge. https://statbel.fgov.be/fr/themes/population/structure-de-la-population#figures (accessed November 15, 2020).

Statistik Austria. 2020. "Population at the Beginning of the Year since 2002 (Regional Status of 2020)." STATcube. https://statcube.at/statistik.at/ext/statcube/jsf/tableView/tableView.xhtml (accessed November 15, 2020).

Statistisk Sentralbyra. 2020. "Population, by Citizenship (M) (UD) 2016–2020." Statbank (Norway). https://www.ssb.no/en/statbank/table/11366/?rxid=d64a36de-8b71-45a4-9ee5-fe3b8462e210 (accessed November 15, 2020).

Statistiska Centralbyran. 2020a. Immigration Statistics (Sweden), provided by e-mail to authors. (accessed September 6, 2020).

Statistiska Centralbyran. 2020b. "Foreign Citizens by Country of Citizenship, Age and Sex. Year 1973–2019." Statistikdatabasen (Sweden). http://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0101__BE0101F/UtlmedbR/?rxid=b83e5bbd-958a-4655-aa40-486ba2ca09a3 (accessed November 15, 2020).

Statistisches Bundesamt. 2020a. "Migration between Germany and Foreign Countries: Germany, Years, Citizenship." Genesis Database. https://www-genesis.destatis.de/genesis//online?operation=table&code=12711-0007&bypass=true&levelindex=1&levelid=1616412990033#abreadcrumb (accessed February 18, 2021).

Statistisches Bundesamt. 2020b. "Foreigners: Administrative Districts, Reference Date, Sex, Country Groups/Citizenship." Genesis Database. https://www-genesis.destatis.de/genesis//online?operation=table&code=12521-0041&bypass=true&levelindex=1&levelid=1616413021684#abreadcrumb (accessed February 18, 2021).

Tilastokeskus. 2020. "Citizenship According to Age and Sex by Region, 1990–2019." Statistics Finland's PxWeb Databases. https://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__vrm__vaerak/

statfin_vaerak_pxt_11rg.px/ (accessed November 15, 2020).

Straubhaar, Thomas. 1988. "International Labour Migration within a Common Market: Some Aspects of EC Experience." *Journal of Common Market Studies* XXVII (1): 45–62.

Von Ehrlich, Maximilian, and Henry G. Overman. 2020. "Place-Based Policies and Spatial Disparities across European Cities." *Journal of Economic Perspectives* 34 (5): 128–49.

Fiscal Policy in Europe: Controversies over Rules, Mutual Insurance, and Centralization

Florin Bilbiie, Tommaso Monacelli, and Roberto Perotti

During the debates in the late 1980s and early 1990s that led to the creation of the euro, controls on national-level budget deficits and debt were among the key objects of contention. Strict limits for deficits and debt were enshrined in the Maastricht Treaty of 1992. Why this seeming fixation with the deficit and debt in a monetary union? The standard account is that countries of northern Europe, with Germany in the lead, worried that some members would run lax fiscal policies and eventually force the other countries to bail them out or force the common central bank to monetize their deficits, thus generating higher inflation—as had happened in the 1970s and 1980s (James 2012). In fact, the Treaty also prohibited the European Union or a member state from "assuming the commitments" of other governments, and the Eurosystem (the combination of the European Central Bank and the national central banks of the eurozone) from "purchasing directly" the debt instruments of a government. This was the famous "no-bailout clause" that, as we shall see, was framed in sufficiently generic terms to allow for a wide range of interventions during the later financial and debt crises of the eurozone.

■ *Florin Bilbiie is Professor of Economics, University of Lausanne, Lausanne, Switzerland. Tommaso Monacelli and Roberto Perotti are Professors of Economics, Bocconi University, Milan, Italy. All three authors are Research Fellows, Centre for Economic Policy Research, London, United Kingdom. Monacelli and Perotti are Fellows, Innocenzo Gasparini Institute for Economic Research (IGIER), Milan, Italy. Perotti is a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are florin. bilbiie@gmail.com, tommaso.monacelli@unibocconi.it, and roberto.perotti@unibocconi.it.*

However, the justification for debt limits shifted in recent years. The European Central Bank purchased enormous quantities of government debt starting in 2015 as part of its quantitative easing policy—more than if it had monetized the deficits of all eurozone countries—and inflation did not budge. Yet budget deficits and government debt have become, if anything, even *more* central to the debate on the reform of the eurozone, at least until the pandemic recession (more on this below). The focus shifted from fears of inflationary pressure to fears of a variety of spillovers and contagion effects stemming from the sovereign default of a monetary union member.

In the rest of the paper, we discuss how the focus of the debate evolved, and the various proposals concerning the use of national and supranational fiscal policy in the context of the European Union and the eurozone.[1] We first follow up on the evolution of rules-based approaches to govern national debt, which have proven largely ineffective. We then turn to proposals for mutual insurance, including pan-European systems of deposit insurance, unemployment insurance, and a pre-agreed "orderly restructuring scheme" for sovereign debt. Finally, we consider proposals for a more centralized European fiscal policy, including policies of debt mutualization and a greater degree of fiscal union. At their core, these arguments over European fiscal policy are a manifestation of two age-old debates: i) rules versus discretion, and ii) risk sharing versus risk reduction and market discipline. An important recent contribution that has tried to reconcile this trade-off is the manifesto of the "7+7" (Bénassy-Quéré et al. 2018), a group of seven French and seven German economists, that has sparked a large and fruitful policy debate on which we draw freely throughout this paper.

To impose some discipline on the discussion, we will take seriously the political constraint whereby large and persistent unidirectional transfers between EU members are politically infeasible. In standard models of insurance across countries, shocks are temporary and randomly distributed, so that expected transfers between countries are typically zero over time. But in the real-life European Union, countries have highly persistent differences in their economies, including different levels and riskiness of their government debts; thus, shocks also tend to be highly persistent. Virtually all policies for risk-sharing arrangements and centralized stabilization face a political challenge, because they are likely to generate either a large, one-off transfer which is unlikely to be offset by a transfer in the opposite direction over a politically realistic horizon (as in the Greek bailout), or long periods of transfers from the core to the periphery, lasting perhaps a decade or a generation (what we call "persistent, unidirectional transfers" for short). Core-country politicians

---

[1] Now that the United Kingdom has brexited, there are 27 European Union (EU) members, all of which are also members of the Economic and Monetary Union. Only 19 of these, however, have adopted the euro and form the eurozone. All countries that joined the European Union after the creation of the eurozone are waiting to qualify for the eurozone; two older members of the European Union, Denmark and Sweden, have chosen not to adopt the euro for the time being.

would have difficulties justifying such transfers to their electorate, especially in the current political atmosphere where nationalist feelings are resurgent everywhere.

In this respect, we recognize that the difference between "core" and "periphery" countries is real: not only are their economies and cultures different in important ways, but the fact of the matter is that a "core eurozone" could survive and maybe prosper, while a "periphery eurozone" is unlikely to be viable or to be of interest to its potential members. Hence, core countries have more bargaining power, and there is nothing to be gained by ignoring this fact. In emphasizing the distinction between different visions of "core" and "periphery" countries, we acknowledge our particular debt to Brunnermeier, James, and Landau (2016).

Unlike many scholars in other social sciences, we do not think that in the long-run a eurozone can be based on anything other than self-interest. In virtually all European countries, nationalistic and Euro-skeptical parties have become more prominent in recent years: but even at the best of times, appeals to notions like "European solidarity" are unlikely to move many voters in the core countries, except perhaps for short, emotionally charged periods. During the Greek crisis in 2010–2015, one of the most popular arguments in the Italian media in support of a European bailout was that Greece was the cradle of the European civilization—the land of Plato and Aristotle—at a time when Germany was barely inhabited. During the Italian debt crisis of 2011–2012, similar arguments were shifted forward by a few centuries to encompass the glories of the Roman empire and of the Italian Renaissance 1,000 years later. We doubt that many German taxpayers and voters found these arguments convincing.

The recent attempts to negotiate an EU fiscal response to COVID-19 illustrated these core-periphery dynamics, but also revealed a degree of added flexibility. In July 2020, the "frugal four" countries—Austria, Finland, the Netherlands, and Sweden—openly defied an agreement by France and Germany to distribute large grants to countries of southern Europe particularly hit by the pandemic, on the ground that these countries had a long history of mismanagement of their public finances. Three of the "frugal four" countries were led by social democratic governments, facing high pressure at home from nationalistic parties. However, by the end of 2020 an agreement was reached to distribute €750 billion borrowed by the union to member countries, based in part on the effects of the pandemic: for instance, Italy got the highest share, €209billion, of which about €80 billion was in grants and the rest in low-interest loans. While our discussion will highlight how efforts toward an EU-wide fiscal policy have been affected by the pandemic, many of the issues described here existed before the pandemic and seem likely to outlast it.

It is also important to note what this paper is *not* about. The European Union has its own yearly budget, amounting to about €160 billion or 1 percent of the total EU gross national income, 85 percent of which is spent on three items: infrastructure projects, structural funds to less developed regions of the European Union, and agriculture. This budget is balanced every year, and it is funded mostly by three items: national contributions proportional to each member's gross national income (65 percent), custom duties (15 percent), and shares in each member's value-added

tax (10 percent). The present paper has nothing to say about this common EU budget. However, in the last section we do discuss the new €750 billion program mentioned above because this is a potential game-changer in the common EU budget.

A Rules-Based Approach

A rules-based approach to fiscal policy requires compliance with specific limits on national government deficits and debt as part of membership in a monetary union. However, there are cases where it may seem wise to bend or set aside such rules, and the list of such exceptions could defang the rule itself. As we will discuss, the pendulum has swung back and forth between tighter and looser fiscal rules. At present, there is a widespread agreement that the attempts to write enforceable fiscal rules with appropriate exceptions are too complicated, unwieldy, even arbitrary. In our view, the quest for alternative rules does not seem to have made substantial progress and might have in fact led to a regress.

The Evolution of European Fiscal Rules

The original Maastricht Treaty of 1992 envisioned two main fiscal policy rules: a limit on annual budget deficits of 3 percent of GDP, and a limit on the accumulated debt/GDP ratio of 60 percent.[2] An exception was allowed only in the case of a downturn of at least 2 percent of GDP. Countries violating these limits could be subject to an Excessive Deficit Procedure, in which the European Commission recommends measures to be taken and monitors the outcomes. By fall 2020, all EU members, except Luxembourg and Sweden, had been subject to at least one Excessive Deficit Procedure. To reinforce these rules, the 1997 Stability and Growth Pact also held that countries should have a budget position "close to balance or in surplus" over a complete business cycle. Still, these limits were honored mostly in the breach: in the 13 years between 1999 and 2011, they were violated by Germany in five years, by Italy in eight years, and by Greece in all years (for the complete list, see Wyplosz 2013).

It was soon recognized that this set of rules was too vague and also dangerous, because it imparted a strong pro-cyclical bias to fiscal policies. When an economy is hit by a negative demand shock, the budget deficit and government debt as shares of GDP rise automatically (because the numerators of both ratios increase and the denominator decreases). Hence, for countries close to the limit, the rules called for a move to a tighter fiscal policy exactly when the economy is hit by a negative shock.

One potential answer is to impose limits to the "structural" deficit instead of the actual deficit, where the "structural deficit" is an estimate of what the budget

---

[2]There are various accounts of where these numbers came from, none of which are verifiable; we just note that in steady state the two numbers are compatible with each other if the rate of growth of nominal GDP is 5 percent, a figure that some countries have not experienced for decades.

deficit would be if the economy were at potential output and one ignored one-off expenditures and revenues. Indeed, a 2005 reform assigned each country a Medium Term Objective, or a target for the structural balance: more lenient if the country had low debt and high potential growth and stricter in the opposite case. A country was allowed to deviate from its Medium Term Objective if it introduced a structural reform, provided it had the effect of raising potential growth and conditional on getting back to the path within four years. Many elements of judgement and uncertainty were involved in all the steps of the process: in estimating *potential* output—a somewhat mysterious and model-dependent object, in estimating the structural deficit, and in assessing whether a structural reform qualified for a deviation from the Medium Term Objective.

After the 2011 debt crisis, the pendulum swung in the opposite direction, with calls by northern European countries to make the rules on government debt more binding. The resulting set of reforms included four steps aimed at a better enforcement of the Stability and Growth pact and two others aimed at other macroeconomic indicators, and thus was nicknamed the "Six Pack." It transformed the deposits in case of noncompliance, as envisioned in the Stability and Growth pact, into fines of up to 0.2 percent of GDP (plus a variable component). Predictably, these fines were never applied. It introduced the "debt brake," by which a country with more than a 60 percent debt-to-GDP ratio would have to reduce it at an average speed of one-twentieth of the excess per year, where the average is calculated over the last three years. On the other hand, the Six Pack relaxed the stringency of the rules by introducing an escape clause from the Medium Term Objective, in case of an "unusual event outside the control of the Member State, which has a major impact on the financial position of the general government, or in periods of severe economic downturn for the euro area or the Union as a whole." To assuage those countries that regarded the policies adopted by Germany as a main source of deflationary pressure on the whole eurozone, the Six Pack also introduced the Main Macroeconomic Imbalance procedure, which would take into account a number of macroeconomic variables, including for instance an excessive current account surplus. But defining an excessive current account surplus is at least as subjective and controversial as defining potential output. Predictably, the Main Macroeconomic Imbalance procedure has been largely toothless as well.

The Fiscal Compact, signed at the end of 2012, incorporated many of these provisions and added more: in particular, still stricter limits for the Medium Term Objectives and several measures to ensure a better enforcement, including an independent "Fiscal Council" in each country and "the obligation to implement measures to correct the deviations over a defined period of time." Once again, it is not clear what this generic wording could achieve in practice. The Fiscal Compact also required countries to enshrine the various fiscal rules (the 3 percent maximum deficit, the Medium Term Objective, the debt brake) into their constitutions. This is widely, but erroneously, interpreted as the requirement of a constitutional balanced-budget amendment.

Table 1
The "Matrix" for Required Annual Fiscal Adjustment

	Condition	*Required annual fiscal adjustment (percentage points of GDP)*	
		Debt ≤60 percent and low/ medium sustainability risks	*Debt >60 percent or high sustainability risks*
Exceptionally bad times	Real growth < 0 or output gap < −4	No adjustment needed	No adjustment needed
Very bad times	−4 ≤ output gap < −3	0	0.25
Bad times	−3 ≤ output gap < −1.5	0 if growth below potential, 0.25 if growth above potential	0.25 if growth below potential, 0.5 if growth above potential
Normal times	−1.5 ≤ output gap < 1.5	0.5	> 0.5
Good times	Output gap ≥ 1.5	> 0.5 if growth below potential, ≥ 0.75 if growth above potential	≥ 0.75 if growth below potential, ≥ 1 if growth above potential

Source: European Commission (2015, p. 20)

In 2015 the pendulum swung back again, reflecting a backlash against the perception that undue fiscal austerity had been imposed. To take into account cyclical conditions, the European Commission devised a matrix (shown in Table 1) specifying the required annual fiscal adjustment towards the Medium Term Objective depending on the growth rate and the debt level of a country. In addition, an array of exceptions and qualifications to the Medium Term Objective itself were introduced, leading to a bewildering cobweb that even experts in the field have trouble navigating. As one example, countries were allowed to deviate temporarily from their Medium Term Objective (or the adjustment path towards it) to accommodate investment, provided that "their GDP growth is negative or GDP remains well below its potential; the deviation does not lead to an excess over the 3 percent deficit reference value and an appropriate safety margin is preserved; investment levels are effectively increased as a result; the deviation is compensated within the timeframe of the Member State's Stability or Convergence Programme" (European Commission 2015 p. 9). As another exception, a country that enacts "structural reforms" may now have its deficit deviate by 0.5 percentage points of GDP from its Medium Term Objective "provided that such reforms (i) are major, (ii) have verifiable direct long-term positive budgetary effects, including by raising potential sustainable growth, and (iii) are fully implemented" (European Commission 2015, p. 12).

The budgetary and growth effects of structural reforms and public investment are largely guesswork, often estimated as the result of a political process of give and take. In 2015, Italy was allowed to deviate from its Medium Term Objective on

the basis of three different clauses: the "unusual events" clause for 0.12 percentage points of GDP (half to cover the costs of the refugee crisis and half for the security crisis after terrorist attacks in Europe); the "investment" clause for 0.25 percentage points; and the "structural reform" clause for 0.50 percentage points. This was based on a submission to the European Commission envisioning that these reforms would raise the level of Italy's GDP by 1.8 percent in 2020 and 7.2 percent in the long run (Italian Ministry of Economy 2015, p. 48)

In a large and diverse union, some exceptions to the rules are inevitable: almost by definition, rules cannot foresee all the relevant contingencies. But the overall result has been unwieldy legislation, endless litigation, backdoor bargaining, and ultimately loss of trust in the European institutions. Any government (and especially any nationalist party) in any country can point to an episode where they can argue that their own country was treated unfairly by their partners according to some criterion.

An Expenditure Rule?

As an alternative way to correct the procyclical bias in rules about debt and deficits, several authors have proposed an expenditure rule (for example, Bénassy-Quéré et al. 2018; Claeys, Darvas, and Leandro 2016; Darvas, Martin, and Ragot 2018; Feld et al. 2018; European Fiscal Board 2019; Mohl and Mourrel 2020). These proposals are all variants of the following approach: the growth rate of government spending net of interest payments, unemployment benefits, one-off expenditures, a smoothed version of public investment, and the discretionary change in tax revenues—call it "adjusted spending growth"—cannot exceed a smoothed version of the growth of potential GDP in nominal terms over a window of ten years or so. Shortfalls of adjusted spending relative to this ceiling can be saved in an accumulation account, whose balance can be spent to finance excesses of adjusted spending in other years. Most proposals also envision a "debt-brake"—that is, in countries with a higher debt/ GDP ratio the expenditure ceiling would be revised downward. Escape clauses and further adjustments for the cyclical conditions are also typically envisioned. All the judgment calls (estimates of nominal potential growth, smoothing of public investment, the adjustments for the debt-to-GDP ratio and cyclical conditions, escape clauses, and so on) would be made by a combination of the European Commission, a European Fiscal Council, and national fiscal councils.

We list three advantages that have been claimed for a so-called expenditure rule. We find all of them highly debatable (see also Barnes and Casey 2019). First, the expenditure rule is supposed to be more transparent than a rule based on the structural balance. However, the discretionary change in tax revenues is conceptually the same as the change in cyclically adjusted tax revenues; hence, conceptually the only difference between adjusted spending and structural balance is that the former subtracts public investment. If the estimate of adjusted expenditure is based on more reliable estimates of the discretionary change in tax revenues, the latter could and should be applied to the estimate of the structural deficit as well. In addition, the construction of an expenditure rule involves *more* steps and many actors,

some of which—like the national fiscal councils in some countries—have not always proved to be authoritative.

Second, the expenditure rule is claimed to be less procyclical than the deficit rules currently in place: in a recession, when GDP grows less than the ceiling, and adjusted expenditure grows at the ceiling, the actual (as opposed to the adjusted) expenditure to GDP ratio increases, and the opposite in periods of high GDP growth. However, as the current rule is specified in terms of the structural, not the actual deficit, it is not clear why this should be the case.

Third, the expenditure rule is supposedly less subject to the vagaries of cyclical adjustment, because revisions to the estimates of expected nominal potential output are less frequent and smaller than revisions of the estimates of potential output. This might be correct if estimating and forecasting the nominal *growth* of potential output over a window of ten years is less uncertain than estimating the *level* of potential output over a horizon of a few years. We do not see a compelling reason for this.

Thus, in our view there is no clear logical or practical advantage of an expenditure rule relative to setting a Medium Term Objective for what is already just a slightly different version of the structural primary balance. In fact, we think it would be a further step backward in terms of transparency with no obvious gains in terms of countercyclicality. Perhaps most surprisingly, a kind of "expenditure benchmark" is already in place among the EU rules: the "Six Pack" of 2011 stipulates that "public spending must not rise faster than medium-term potential GDP growth, unless it is matched by adequate revenue increases." In implementing this rule, the construction of "public spending" is essentially the same in the expenditure rule proposals we have described above (on the performance of the expenditure benchmark, see Barnes and Casey 2019; Mohl and Mourre 2020).

In another recent effort to improve the current institutional setup, Blanchard, Leandro, and Zettelmeyer (2021, p. 20, including quotation from Sunstein 1995) argue that because rules cannot foresee all possible contingencies, they should be replaced by "standards" that "leave room to accommodate the particulars of individual circumstances [. . .] Rules have the advantage of providing greater clarity ex ante. But a case-by-case approach may be preferable when 'public authorities cannot design general rules, because they lack relevant information . . . or rules [would] be poorly suited to new circumstances turned up by unanticipated developments.'" As an example of a standard, they cite the current article 126 of the EU Treaty; "Member states shall avoid excessive government deficits." They propose that this should be made operational by resorting to a debt sustainability analysis. Whether standards of this type can be a practical basis for a yearly political process involving 27 different countries is very much an open issue: for one thing, debt sustainability analyses are probably even more contentious and subject to large areas of uncertainty than, say, the estimation of potential output.

During the 2020 pandemic, the entire framework of European fiscal rules has been *de facto* suspended: the "general escape clause" has become operational, allowing member states to depart from the adjustment path towards their Medium Term Objective. This step is meant to be temporary, but there is a widespread

perception that a radical revision of the whole framework will be needed once the emergency is over. It is most likely, however, that *some* rules and limits will remain as the cornerstone of the framework.

Mutual Insurance

We define "mutual insurance" as inter-country transfers that occur in response to asymmetric shocks. These transfers can occur automatically, via a facility like European unemployment insurance, or on an ad-hoc basis. There are obviously several market failures that could justify mutual insurance between governments. We emphasize that in a currency union with nominal rigidities, even with complete financial markets, member countries might hold a suboptimal degree of insurance. The reason is that, with a fixed nominal exchange rate, international real relative prices do not adjust efficiently to asymmetric disturbances. In this context, it is constrained efficient from a union-wide perspective to resort either to variations in government spending (Galí and Monacelli 2008) or to cross-country transfers (Farhi and Werning 2017) to stabilize the economy. In practice, however, difficulties arise in implementing mutual insurance between countries, mainly because almost any conceivable implementation of such schemes in the current eurozone would give rise to large, persistent unidirectional transfers.

We focus on two candidates for mutual insurance that are closely related to fiscal policy: a European deposit insurance scheme and a European unemployment insurance scheme. Both have figured prominently in the academic and policy debate; both have failed to make much inroad among policymakers. We also discuss a European orderly resolution scheme for government debt, which could reassure core countries that mutual insurance is less likely to lead to bailouts of periphery countries.

A European Deposit Insurance Scheme and Safe Assets

Europe has largely completed the first two steps of a banking union: a common bank supervision and a common procedure for resolving failed banks. However, the third step, a European deposit insurance system, has faced the opposition of Germany and other core countries.[3]

At an abstract level, one can think of a banking union as a risk-sharing tool because its ultimate goal is to ensure that risk-adjusted funding costs for banks are the same across member countries, and in particular that these risk-adjusted funding costs are independent of sovereign risk.[4] Like all instances of insurance, a European deposit insurance system raises an issue of moral hazard that should

---

[3] The German Finance Minister has made some recent overtures to a European deposit insurance system (Sholz 2019).

[4] Martinez, Philippon, and Sihvonen (2019) study the comparative properties of risk sharing via a banking union or a capital market union.

not be underestimated: that is, governments backed by a European deposit insurance system and facing a crisis could pressure their banking system to purchase their domestic government debt and to keep lending to firms in distress. In this way, a European deposit insurance system raises the danger of a "doom loop"—the vicious circle of negative feedback that can emerge between a country's debt and its financial sector, which could lead to large and persistent transfers to periphery countries.[5] The "doom loop" starts with the home bias of banks in their holdings of government bonds. This increases the probability of a "bad" (or "sunspot") equilibrium, in which the government is expected to bail out banks in distress, and banks are in distress because the expectation of a government bailout reduces the value of domestic government bonds on their balance sheets. Brunnermeier et al. (2017, especially its online Appendix) provide a model that formalizes these ideas.

The doom loop first became the object of a heated debate at the time of the sovereign debt crisis of 2011–2012, when Italian and Spanish banks surprised many by using the large injection of liquidity by the European Central Bank to double down on their holdings of risky domestic sovereign debt. In a currency union, where governments have neither access to their own "printer of last resort" nor the option to devalue their exchange rate, the spillover effects on other countries of a financial system breakdown coupled with a confidence crisis in the government debt could be particularly disruptive. Greece represents only 3 percent of the eurozone GDP, yet for several years after 2010 talks about a possible Greek default kept the entire eurozone on edge and fueled a raging debate about a possible euro breakup.

To minimize the moral hazard problem and address the "doom loop" issue, four solutions are typically put forward, all revolving about the notion of reducing the home bias of the banking sector in countries with large and risky government debt. A first proposal envisions adjusting the deposit insurance premia for bank-specific and country-specific risk, along with a first loss to be taken by the domestic government. Negotiating the appropriate premia would be challenging in practice, as they must be based on measures of the solidity of each country's banking system that are likely to be even more contentious than government debt ratings or country-specific potential outputs.

A second idea involves "concentration charges," in which the higher the share of a sovereign in the total assets of a bank (or in the total sovereign holdings), the higher the capital charge coefficient applied to the holdings of that sovereign in that bank (for example, German Council of Economic Experts 2015; Veron 2017; and the proposal by the German Finance Minister, Sholz 2019). Unsurprisingly, the mention of concentration charges is anathema to periphery governments with high public debt and to their central bankers.

---

[5] Carmassi et al. (2020) argue that regardless of which several different risk-weights are used, the cross-subsidization via a European deposit insurance system would be zero or minimal even in the presence of bank failures, well in excess of those seen in the Great Recession. However, their results are based on symmetric shocks (in each country, banks representing a given percentage of that banking system's assets fail, and the loss rate is the same across countries).

A third proposal envisions differentiated capital charges depending on the risk of each sovereign debt. The European Union has chosen to make use of the discretion allowed by the Basel rules and has adopted a zero-risk weight on the government debt of any EU country and held by any EU bank under the "standardized approach" (for a useful summary of the issues, see European Parliament 2019). This proposal too is anathema to high-debt periphery countries. In addition, it would largely refute current policy of the European Central Bank, which applies a rather crude differentiation to the haircut of government bonds in its repo operations, with *de facto* only two categories of sovereign risk. However, the simulations of Alogoskoufis and Langfield (2019) show that even following the finer ratings of rating agencies would not achieve a substantial reduction in risk because these ratings are noisy and unreliable.

A fourth proposal is the creation of a safe asset, often called European safe bonds or "Esbies," via financial intermediation. The idea is that if periphery banks had access to a large supply of well-diversified safe assets, they would be somewhat insulated against a loss of confidence in their own home-country sovereign debt. To create Esbies, Brunnermeier et al. (2011) and Brunnermeier et al. (2017) propose pooling eurozone sovereign debt according to their GDP weights, and then to divide this pool of debt into tranches: the junior tranche would be the first to take any losses, which means that the senior tranche—the Esbies—can be made as safe as one wishes. For example, simulations in Brunnermeier et al. (2017) show that when the junior tranche absorbs the first 30 percent of losses, and under the observed matrix of correlations, the senior tranche can effectively be made as safe as euro-denominated bonds issued by Germany's government. Note that the pooling and tranching of European sovereign debt could be done by the market once the European Central Bank announces that it will accept the senior tranche as collateral.[6] The European Parliament recommended the adoption of Esbies in April 2019, and Garicano (2019) proposes a path to the implementation of Esbies.

For present purposes, the key insight is that European safe bonds could reduce the risk of a "doom loop" by limiting the home bias of banks, in turn reducing the probability of the "bad" equilibrium described above. As a result, governments of core countries would have less reason to fear that a European deposit insurance scheme would become a vehicle for large unilateral transfers.

---

[6] Several variants of the notion of Esbies have been proposed. There is also a large theoretical literature on safe assets, in which a general theme is that the supply of safe assets has shrunk dramatically in recent years (for discussion in this journal, see Caballero, Farhi, and Gourinchas 2017). The "safety trap" of Caballero and Farhi (2017) and Caballero, Farhi, and Gourinchas (2020), as summarized in the model of Caballero, Farhi, and Gourinchas (2016), suggests that an excess demand for safe assets can lead to a recession. Essentially, think of a natural (real) safe interest rate, which equates demand and supply of safe assets at the natural level of output. If a central bank cannot reduce the actual (real) safe interest rate to this level, perhaps because it faces a lower bound at (or near) zero percent in setting this rate, then the equilibrium is obtained via a recession that reduces the demand of safe assets (savings) until it is equal to supply.

The key question is: if European safe bonds were available, would banks in periphery economies buy them willingly? Answering this question requires studying the reasons why financial institutions of periphery countries have shown a strong home bias for purchasing debt issued by their own national government. Although there is no consensus, we list four—not necessarily mutually exclusive—reasons for this home bias.

First, home-country governments can influence the investment decision of financial institutions via moral suasion (for example, Becker and Ivashina 2018; Acharya and Steffen 2015; De Marco and Macchiavelli 2010; Horvath, Huizinga, and Ioannidou 2019).

Second, holding troubled home-country debt can be part of a "gambling for resurrection" strategy by a distressed bank. After a deep recession, a bank may find itself with a large share of non-performing loans. Risky sovereign debt has high yields, while bank regulators treat it as non-risky. Thus, banks at the periphery have loaded up on domestic sovereign debt in times of economic stress—and doubled down on this strategy when the opportunity arose, such as with the two large Long Term Refinancing operations of the European Central Bank in 2011–2012.

Third, the marginal cost of borrowing for banks (the interest paid on the bonds they issue) is closely related to the interest paid by their sovereign—and this is largely independent of the amount of sovereign debt they hold. Credit rating agencies typically do not rate the debt of financial institutions more than two notches above their home country sovereign debt. In fact, Constâncio (2018) shows that the increase in riskiness of Italian banks during the debt crisis of 2011–2012 (as measured by premia on the related credit default swaps) was unrelated to their individual exposure to home country sovereign debt.

Fourth, life insurance companies also hold large amounts of sovereign debt. Typically, the guaranteed return on their life insurance contracts are linked to the *domestic* interest rate, which means that insurance companies of the periphery have an incentive to invest their assets disproportionately in home country sovereign debt.

If these explanations for the home bias of banks and financial institutions are correct, it is unlikely that there would be a large demand for European safe bonds by the banking sector in the periphery. After all, periphery banks *could* have invested in a safe, euro-denominated asset virtually identical to Esbies—the euro-denominated bonds issued by the Germany called Bund—but they *chose* not to.

More generally, some argue that reducing the home bias of the periphery's banking system is misguided in the first place (for example, Tabellini 2017, 2018). Conditional on periphery countries having high public debt, there are situations in which the home bias in bank holdings of sovereign debt is not only inevitable for the political and economic reasons that we have seen, but is also desirable. As one example, suppose that the "bad" or "sunspot" equilibrium arises from a pure liquidity crisis on government debt, and thus is not associated with a bailout of home-country banks. In this setting, by investing in home-country sovereign debt (perhaps using also cheap liquidity from the Eurosystem), the banking systems of

high-debt countries stabilize the markets of their sovereigns and avoid a debt run. In this view, the original sin is the high government debt: until it is reduced, forcing a reduction of the home bias could be counterproductive. This view also suggests that in order to reduce the high government debt, rules constraining national fiscal policies should be strengthened, not weakened.

For similar reasons, the other proposed solutions to the "doom loop" (like concentration charges and risk weights) could backfire, thus reinforcing rather than weakening the "doom loop." If the exposure of periphery banks to home-country sovereign debt is inelastic to the risk of that debt— perhaps because it arises for reasons of moral suasion by home country governments or gambling for resurrection as mentioned earlier—then concentration charges and risk weights would make domestic banks *more* risky when the sovereign becomes more risky.

A European Unemployment Insurance Scheme

Unemployment insurance is perhaps the most basic and natural case of risk-sharing arrangement, yet a European unemployment insurance scheme has not found much favor with core governments. The reason, once again, is the potential for large and persistent unidirectional transfers. Table 2 displays how much higher unemployment rates have consistently been in the last decade in Italy and Spain, compared with unemployment rates of the "core" eurozone countries (the table also displays the year of the highest difference, 2014 in Italy and 2013 in Spain). In this setting, any European unemployment insurance scheme tied to unemployment rates would have generated enormous and persistent transfers.

How might this prospect be avoided? One could try to design a European unemployment insurance scheme in terms of *differences* of the unemployment rate from a "neutral" or "long-term" level, but this would entail all the analytical and political controversies associated with estimates of notions like "potential output," "potential growth," or the "natural rate of unemployment." Alternatively, a European unemployment insurance scheme could be based on *changes* in unemployment instead of levels, but then one could have large transfers from countries with high yet temporarily decreasing unemployment to countries with low yet temporarily increasing unemployment, which would be unacceptable to the public. Delegating everything to an independent council will not work either because politics can *and will* take over when a large shock makes the stakes high.

Other potential features would either largely defeat the purpose of a European unemployment insurance scheme, or be also difficult to apply in practice. For example, one can imagine adjusting unemployment-insurance contributions for country-level risk; in practice, this would require assessing the risk of a country's unemployment, again a contentious proposal. Limiting the European unemployment insurance scheme to a catastrophic insurance scheme that pays only in the case of extremely large increases in unemployment, would severely limit the usefulness of the scheme; and once again, delegating to an independent body the determination of the catastrophe clause trigger is unlikely to work when it matters most.

Table 2

Difference between the Unemployment Rate of Core Countries and That of Italy and Spain

Country	Difference with Italy			Difference with Spain		
	2009	*2014*	*2019*	*2009*	*2013*	*2019*
Belgium	0.1%	4.2%	4.6%	11.6%	17.7%	8.7%
Germany	1.4%	7.7%	6.8%	12.9%	20.9%	10.9%
France	−0.9%	2.4%	1.5%	10.6%	15.8%	5.6%

Source: European Commission, AMECO Database
Note: The cells of the table show the result of subtracting the unemployment rates row country from the column country.

One might think that Europe could just copy the US system. In the United States, unemployment insurance is a state responsibility, but the federal government has supplemented it in every recession since 1950 (what follows is based on Porter 2021). In 1970, the federal "extended-benefit" program was created, adding federal funding for additional weeks of benefits in deep recessions, but states still had to pay half of the extension. Perhaps because of the hidden moral hazard in this pattern of repeated federal interventions, in 2019 only 29 percent of unemployed workers received a benefit; the percentages were as low as 9 and 11 percent in North Carolina and Florida, respectively, reaching a maximum of 60 percent in New Jersey. The average benefit was one-third of the last wage. During the Great Recession, the Obama administration spent lavishly to fund the unemployment system, but states moved the opposite way. Having exhausted their unemployment insurance funds, they cut benefits, and rather than increasing taxes, they went deeply into debt with the US Treasury ($42 billion by 2011). The political wrangling over the federal extension of benefits during the pandemic further illustrates the tensions at the core of this system.

In a country with high labor and capital mobility, tax competition between states makes it difficult to fund a state unemployment insurance system. On the other hand, federal intervention seems to have created the familiar moral hazard problem. Europe has less labor mobility than the United States, but it is difficult to escape the conclusion that a European unemployment insurance scheme would have to be funded and run at the central level with the problems highlighted above.

A European Orderly Restructuring Scheme

Sovereign defaults are typically chaotic events. This generates a vicious cycle. Exactly because defaults are often chaotic, it might be rational for core countries to limit the contagion and spillover effects by bailing out the defaulting government after the events have occurred. Knowing this, high-debt countries often procrastinate and delay the adjustment; the disruptions that follow reduce the recoverable value when the default becomes unavoidable and may in some cases cause a deep

recession. Panizza (2013a, b) discusses this and other costs of sovereign debt default. For all these reasons, core countries are unwilling to enter a risk-sharing arrangement with periphery countries with high default risk. An orderly restructuring scheme is designed to break this vicious cycle.

An orderly restructuring scheme involves a predictable and orderly process. Two key elements are a predetermined process of restructuring and rescheduling of the debt when a default occurs and a "bail-in" of private creditors: that is, knowing in advance that private creditors too will have to take a loss, and by how much. This increases the recoverable value when default is inevitable, making core countries more willing to enter a risk-sharing arrangement with high debt countries: defaults are more frequent, but they are accompanied by orderly restructuring and a bail-in limiting the losses to the government, rather than a chaotic bailout (for a model in support of this argument, see Gourinchas, Martin, and Messer 2020). In turn, the risk-sharing arrangement makes a European orderly restructuring scheme more acceptable to periphery countries because it absorbs parts of the costs associated with a default *cum* restructuring. The key point of the "7+7 proposal" is precisely that there is complementarity between risk sharing and an orderly restructuring scheme.

In practice, would an orderly restructuring process increase recoverable value? Although we have scant evidence, we know what happened after the introduction of mandatory "collective action clauses" on government bonds issued by eurozone countries with maturities above one year, starting in 2013. This kind of clause allows a specified supermajority of bondholders to agree to a debt restructuring plan that is binding on all bondholders, thus reducing the "holdout" problem. Collective action clauses can therefore be thought of as modest version of a fuller European orderly restructuring scheme, in the sense that they seek to make a restructuring with a bail-in component more likely but less costly. Indeed, sovereign borrowing costs decreased with collective action clauses, which seems to suggest that the costs of default did decline.[7]

A related problem can arise if a government in distress decides to borrow a large incremental sum, because in the absence of a well-defined seniority structure, additional borrowing when close to default hurts all existing creditors. This is different from the case of corporations, where a better-defined seniority structure protects the more senior creditors. As part of an orderly restructuring scheme, the 7+7 group advocates requiring countries to issue junior debt when their debt exceeds a certain threshold: this will increase market discipline, as the country must pay a higher interest rate at the margin.

---

[7] Tabellini (2018) offers the opposite interpretation of the same piece of evidence. In his view, collective action clauses were meant to make default (with bail-in) more likely, hence they should have increased borrowing costs; instead, borrowing costs decreased because holders of government bonds issued under international law enjoy more protection than holders of government bonds issued under national law (one possible reason is that national courts are captive to the domestic government). Tabellini (2018) concludes that, with all of these factors taken into account, issuing debt under international law makes debt renegotiation more difficult, not less.

Critics of proposals for an orderly restructuring scheme raise several concerns. For some, a restructuring of a periphery country's sovereign debt is likely to be at best a partial fix: it will almost surely take down the banking system of that country anyway. Most proposals for a restructuring scheme recognize this problem, although they usually rely on generic recommendations that an orderly restructuring scheme should be accompanied by a reduction of non-performing loans and of the banking system's exposure to domestic sovereigns.

In addition, a restructuring process that forces governments to issue junior debt might backfire because a default on even junior debt would most likely trigger a run on the whole debt stock.[8] But perhaps the main concern is that merely talking about the possibility of a restructuring increases the cost of borrowing in periphery countries and, by increasing the costs of rolling over debt, it might transform a liquidity problem into a solvency problem (for example, De Grauwe and Ji 2018). However, as we have discussed above, if the alternative to an orderly restructuring scheme (and an associated bail-in of creditors) has a high chance of disorderly default, restructuring causes sovereign borrowing costs to decrease instead. Of course, this second outcome assumes that the no-bailout clause is credible: this is precisely what an orderly restructuring scheme with risk sharing is designed to do.

For the critics, proposals for a European orderly restructuring scheme tip the scale too much in favor of market discipline. Instead, they argue, reducing debt in high-debt countries must be achieved before a European orderly restructuring scheme can be introduced. Thus, they tend to argue that rules constraining national fiscal policies should be strengthened, not weakened. This is perhaps the key message of the famous "non-paper" circulated by the then-German minister of finance, Schäuble (2017), which with some oversimplification can be summarized as follows: "Put your house in order by reducing your debt, introduce an orderly restructuring scheme, and then we might talk about a mutual insurance scheme."

Centralized Fiscal Policy, Aggregate Stabilization, Debt Mutualization, and Fiscal Union

The advantage of automatic risk-sharing mechanisms, like deposit or unemployment insurance, is that they work in the background and do not need an explicit

---

[8] Some argue that a better way to introduce a seniority structure in government debt would be to introduce GDP-linked bonds (Tabellini 2017). A GDP-linked bond is *de facto* junior because it pays less in bad times. The idea of GDP-linked bonds goes back to Shiller (1993); for an exposition of key issues, see Cecchetti and Schoenholtz (2017) and Shiller et al. (2018). To date, no country has issued GDP-linked bonds. The leading explanations are the risk premium demanded by investors to take on a more volatile returns, and the liquidity premium associated with a new instrument. For these reasons, Blanchard, Mauro, and Acalin (2016) argue that GDP-linked bonds are more appropriate for countries with high, but not "catastrophically" high, debt. Kim and Ostry (forthcoming) argue that the advantages of GDP-linked bonds have to be set against possible moral hazard considerations, but we find it implausible that governments would induce a recession in order to reduce the value of the principal or interest of their debt.

political agreement every time that they redistribute resources. The limitation is that they have a constrained redistributive capacity for those circumstances in which more risk-sharing might be needed. In these cases, a supranational fiscal authority that implements risk sharing after a crisis might be called for. Such authority might also be needed to respond effectively to aggregate shocks: in the presence of spillover effects, the response by individual countries acting in isolation might be suboptimal. We refer to this role of a supranational fiscal authority as "stabilization policy," to be distinguished from the "risk-sharing" policies discussed above. Finally, a supranational fiscal authority could be the enabler of large pan-European infrastructure investment projects, again in the presence of large spillover effects and coordination problems.

Many proposals for European fiscal policy push in the general direction of "more fiscal policy at the European level." However, this expression can mean a myriad of policies and institutional arrangements that are very rarely spelled out precisely. In what follows, we try to give a sense of the complexities that can arise when trying to give concrete content to the expression.

A European Monetary Fund

In some circumstances, there might be a consensus that a highly focused intervention is needed in a country hit by a particularly negative shock that cannot be addressed by standard risk-sharing arrangements like deposit or unemployment insurance. The European Stability Mechanism is meant to work as a sort of regional-level International Monetary Fund. It leverages a relatively small paid-in capital of €80 billion (paid roughly in proportion to the GDP shares of each country) to borrow on the market. Because it does not borrow more than the total callable capital of a few core countries (about €500 billion), effectively its debt is rated AAA; it then lends to illiquid countries at a rate below their borrowing rates.

Core countries have insisted on three key features. First, to comply with a no-bailout rule, the European Stability Mechanism can lend only to countries whose debt has been deemed "sustainable." Second, by implication, countries with unsustainable debt can borrow only if they restructure their debt. Third, lending via the European Stability Mechanism is subject to various degrees of conditionality, depending on the specific program chosen. Periphery countries object to all three features. The mere possibility of debt restructuring is unacceptable to any periphery government already struggling with the market's perception of its solvency. A periphery government that borrows from the European Stability Mechanism would immediately be accused by a large share of the electorate to yield to austerity plans imposed from outside, although this perception would probably be unfairly polluted also by the reminiscence of the largest intervention by the European Stability Mechanism and its previous incarnation, the second and third Greek programs between 2012 and 2018, which occurred under dramatic circumstances and a stronger conditionality.

In short, core countries want to make the European Stability Mechanism the instrument of risk prevention and the guardian of rules; periphery countries would

like to make it an instrument for stronger risk-sharing, with more resources and fewer conditions attached. Perhaps because of these conflicting views, the European Stability Mechanism has not been used since the Great Recession of 2008–2009. As we write, the two countries most affected by the pandemic, Italy and Spain, appear inclined to reject COVID-related loans at near-zero rates to be spent specifically on the public health system because of the domestic political implications mentioned above, and also because they can still borrow in the market at rates that are barely above those available from the European Stability Mechanism.

Fiscal Union

Some proposals advocate a "fiscal union," by which they seem to mean a centralized European entity with an autonomous taxing and spending authority. This European "finance minister" could spend its resources either as block grants to individual countries or directly on specific projects (like a new high-speed train), or via automatic programs (like a European pension system). It could be funded both by its own dedicated taxes (some suggested examples include an excise tax on single-use plastics or a tax on revenues of digital companies), by shares of tax revenues collected by the member states, or by the issuance of its own debt. Thus, there is a large number of possible combinations of funding and spending patterns, which are rarely specified.

Furthermore, as discussed above, this fiscal union could be an additional instrument for risk-sharing, for stabilization policies, or for coordinating large public investment projects. Supporters of a European finance minister also often argue that it is a precondition for a closer political union.

Core countries sometimes pay lip service to more fiscal integration, but they are largely unenthusiastic about it. Once again, the key problem is its distributional implications. In principle, a fiscal union could be implemented in a distributionally neutral way in the long run: it is easy to imagine a scheme whereby a centralized fiscal authority makes unconditional transfers to member countries, and in the long run the recipients pay back what they receive in present value terms. If one goes beyond pure transfers, however, it is easy to imagine that in practice an expanded remit of a centralized fiscal authority would lead to large and persistent flows of resources from the core to the periphery. The bulk of government spending is on pensions, government employment, health, and other purchases of goods and services; the levels of these expenditures and the systems governing them are very different across European countries. Core countries fear that centralized spending on these items will inevitably flow disproportionately to the higher spending countries. If, to avoid this outcome, more homogenization of policies is imposed as a prerequisite of more centralization, this by definition means going against the collective preferences of some or all member countries on sensitive types of spending. Such an attempt would generate a strong political backlash and could well endanger the union rather than strengthening it (Alesina and Perotti 1998).

These tradeoffs remain unresolved or even unacknowledged in virtually all proposals for more centralized fiscal policy. Many academic economists and

policymakers seem to take it as given that more centralized European fiscal policy is a prerequisite for the survival of the Union and of the eurozone. Without specific details on what this means in practice, however, it is difficult to evaluate this claim.

Eurobonds

If centralized fiscal policy is financed by debt issued at the central level, this step would add another channel of potential redistribution. Currently, any debt issued by the European Stability Mechanism is covered by a *proportionate* guarantee of the member states: in case of a default by the borrower, the other countries will be called to cover the shortfall in proportion to their shares of capital, hence approximately in proportion to their GDP. Others go a step further and propose "eurobonds," a debt issued at the European level covered by a joint and several guarantee of each member country. In a joint and several guarantee, each guarantor can be called upon to pay for the *whole* guaranteed amount in case of default by one or more of the joint issuers. That guarantor can then follow up by asking the other guarantors to contribute their shares. Obviously, eurobonds impose more risk on Germany than on Greece. In many cases, proposals for eurobonds appear to be an *intended* explicit mechanism for planned redistribution, even though exactly how the proceeds of a eurobond issue are distributed to and repaid by the individual countries is almost never specified.[9]

It should come as no surprise that eurobonds, in all their shapes, have been proposed mostly by periphery countries. The debate about eurobonds has been marked by considerable ambiguity and much political posturing: the term is often used to denote any debt issued at the European level, even without joint and several guarantees. This confusion has occurred frequently when referring to the "coronabonds" that will finance the "Next Generation EU" scheme in response to the pandemic recession, to which we now turn.

The Coronavirus and European Fiscal Policy

The pandemic recession that began in 2020 has altered the discussion about European fiscal policy. Arguments over the appropriate rules for limiting debts and deficits have been shut down until later. Issues of a European deposit insurance system or an orderly resolution system have been pushed to the back burner, as well. There has been no movement toward a proper unemployment insurance scheme, but the European Union did enact the SURE program ("Support to mitigate Unemployment Risks in an Emergency"), providing loans up to €100 billion to

---

[9] Over time, a large number of proposals have sought to mitigate the cruder aspects of Eurobonds. We do not have space to review these alternative proposals; suffice it to say that none of the proposals that maintain the joint and several guarantee in some form or proportion have managed to make inroads in core countries.

supplement national expenditure on short-term working schemes. The take-up has been exceptional, and all funds have been loaned out in a few months.

Much of the policy focus during the pandemic has been on a centralized European fiscal response. The road has been a bumpy one, but eventually the European Union agreed on a set of programs, collectively called the Next Generation EU, for €750 billion in total (about 5 percent of the EU GDP), of which the largest component is the Recovery and Resilience Facility for €670 billion. Of this total, slightly more than half will be in the form of loans and the rest in the form of grants. All funds will have to be spent between 2021 and 2026. The European Union will borrow the entire amount and these "coronabonds" will be repaid by 2058. The loans will then be reimbursed by the individual recipient countries, while the part corresponding to the grants will be repaid with new own resources of the European Union (like a digital tax, a carbon border adjustment mechanism, or a financial transaction tax). Notice that it is easy for the public and many politicians of the countries receiving the grants to operate under an optical illusion: obviously not all the grants are a net gift to the recipient, as the latter will have to contribute to the repayment of the whole pool of grants, roughly in proportion to its own GDP.

Is all this a game changer? It *is* the first time that the European Union borrows directly as such; before this event, it was a matter of dispute whether borrowing by the European Union is legal at all. It *is* a fairly large amount. It *is* highly redistributive: Italy, the biggest recipient, will get almost 30 percent of the Next Generation EU funds, almost three times its share in the EU GDP. On the other hand, and unlike what many think or say, the coronabonds are *not* eurobonds with joint and several liability. The biggest recipients, like Portugal, Spain, and possibly Italy and France have already signaled that they will take up the grants (no surprise here), but probably not the loan part.

The key question arises: is this a one-off or will it lead to permanent institutional changes? Of course, it is unlikely that such large grants will be repeated any time soon, but will this experience lead to a new mutual understanding of an expanded role for the European Union as such, relative to national governments in fiscal policy matters? Will it lead to a stronger role of centralized fiscal policy, however defined, and a corresponding downplaying of fiscal rules? Many commentators are convinced that this is the case, that Europe has experienced a "Hamiltonian moment," named after the role of Alexander Hamilton in acting to federalize the state-level debts that had been incurred in US Revolutionary War and its aftermath. These commoners seem to envisage permanent Hamiltonian effects: a first step towards building a true European fiscal capacity, a stepping-stone for a sizeable future expansion of the common EU budget. We are less sure.

The new EU debt facility consists of transfers, with strings attached as to how they can be spent (mostly on digitalization, "green transition," and infrastructure projects). Making these transfers permanent and preserving their highly redistributive bias to the periphery countries would be politically unfeasible for the reasons that we have mentioned throughout this paper. Making them distributionally neutral (such that the transfers are equal to the present value of the resources paid

back by the recipient) would make little sense, as long as all countries involved still have access to capital markets, as they do now: the only advantage to periphery countries would be a small savings in interest payments as the European Union can borrow at slightly lower rates than its periphery members. Moving from transfers to direct spending by the center would involve resolving the large diversity of collective preferences on fiscal policy and would require a wholly different institutional setup with a true central fiscal authority.

For some, a true game changer is the new autonomous taxing authority of the European Union, which will collect its own taxes to repay the coronabonds issued to finance the grants. This development is conceptually a new one: in the past, the European Union essentially relied on shares in national VAT or contributions by member countries to fund its modest budget. However, repaying €310 billion over a horizon of years implies increasing the taxing capacity of the European Union by about €11 billion per year on average (depending on the maturity profile of the debt), which is less than 0.1 percent of the EU GDP. This does not seem to be a quantum leap.

As we mentioned, several core countries only grudgingly agreed to the Next Generation EU program. It is true that, unlike the funds disbursed by the common EU budget so far (so-called Structural Funds), the facility makes the disbursement contingent on meeting certain pre-agreed criteria. But evaluating the effects of the large transfers to periphery countries will be difficult in itself, and should the perception spread among the electorate of core countries that these funds have not been spent productively by the recipients, it is easy to imagine a backlash against, and not more support for, any form of centralized fiscal policy and increased mutual insurance.

Thus, although additional European-wide fiscal policy steps may well be taken in the years ahead, the policy debate and political negotiations will continue to largely revolve around the key issues we have highlighted in this paper.

References

Acharya, Viral V., and Sascha Steffen. 2015. "The 'Greatest' Carry Trade Ever? Understanding Eurozone Bank Risks." *Journal of Financial Economics* 115 (2): 215–36.

Alesina, Alberto, and Roberto Perotti. 1998. "Economic Risk and Political Risk in Fiscal Unions." *Economic* Journal 108: 989–1009.

Alogoskoufis, Spyros, and Sam Langfield. 2019. "Regulating the Doom Loop." European Central Bank Working Paper 2313.

Barnes, Sebastian, and Eddie Casey. 2019. *Is the Expenditure Benchmark up to Standard?* Dublin: Irish Fiscal Advisory Council.

Becker, Bo, and Victoria Ivashina. 2018. "Financial Repression in the European Sovereign Debt Crisis." *Review of Finance* 22 (1): 83–115.

Bénassy-Quéré, Agnès, Markus Brunnermeier, Henrik Enderlein, Emmanuel Farhi, Marcel Fratzscher, Clemens Fuest, Pierre-Olivier Gourinchas et al. 2018. "Reconciling Risk Sharing with Market Discipline: A Constructive Approach to Euro Area Reform." CEPR Policy Insight 91.

Blanchard, Olivier, Alvaro Leandro, Jeromin Zettelmeyer. 2021. *Redesigning EU fiscal rules: From rules to Standards.* German Federal Ministry of Finance.

Blanchard, Olivier, Paolo Mauro, and Julien Acalin. 2016. "The Case for Growth Indexed Bonds in Advanced Economies." Peterson Institute Policy Brief PB16–2.

Brunnermeier, Markus K., Luis Garicano, Philip R. Lane, Marco Pagano, Ricardo Reis, Tano Santos, David Thesmar, Stijn Van Nieuwerburgh, and Dimitri Vayanos. 2011. "European Safe Bonds (ESBies)." https://personal.lse.ac.uk/vayanos/Euronomics/ESBies.pdf.

Brunnermeier, Markus K., Harold James, and Jean-Pierre Landau. 2016. *The Euro and the Battle of Ideas.* Princeton: Princeton University Press.

Brunnermeier, Markus K., Sam Langfield, Marco Pagano, Ricardo Reis, Stijn Van Nieuwerburgh, and Dimitri Vayanos. 2017. "ESBies: Safety in the Tranches." *Economic Policy* 32 (90): 175–219.

Caballero, Ricardo J., and Emmanuel Farhi. 2017. "The Safety Trap." *The Review of Economic Studies* 85 (1): 223–74.

Caballero, Ricardo J., Emmanuel Farhi, and Pierre-Olivier Gourinchas. 2016. "Safe Asset Scarcity and Aggregate Demand." *American Economic Review: Papers & Proceedings* 106 (5): 513–18.

Caballero, Ricardo J., Emmanuel Farhi, and Pierre-Olivier Gourinchas. 2017. "The Safe Asset Conundrum." *Journal of Economic Perspectives* 31 (3): 29–46.

Caballero, Ricardo J., Emmanuel Farhi, and Pierre-Olivier Gourinchas. 2020. "Global Imbalances and Policy Wars at the Zero Lower Bound." CEPR Discussion Paper DP14424.

Carmassi, Jacopo, Sonja Dobkowitz, Johanne Evrard, Laura Parisi, André Silva, and Michael Wedow. 2020. "Completing the Banking Union with a European Deposit Insurance Scheme: Who Is Afraid of Cross-Subsidisation?" *Economic Policy* 35 (101): 41–95.

Cecchetti, Stephen, and Kim Schoenholtz. 2017. "GDP-linked Bonds: A Primer." *VoxEu*, March 17. https://voxeu.org/article/gdp-linked-bonds-primer.

Claeys Grégory, Zsolt M. Darvas, and Álvaro Leandro. 2016. "A Proposal to Revive the European Fiscal Framework." Bruegel Policy Contribution 2016/07.

Constâncio, Vítor. 2018. "Completing the Odyssean Journey of the European Monetary Union." Remarks, ECB Colloquium on the Future of Central Banking, Frankfurt am Main, May 16–17, 2018.

Darvas, Zsolt, Philippe Martin, and Xavier Ragot. 2018. *European Fiscal Rules Require a Major Overhaul.* French Council of Economic Analysis.

De Grauwe, Paul, and Yuemei Ji. 2018. "Financial Engineering Will Not Stabilize an Unstable Euro Area." VoxEu, March 19. https://voxeu.org/article/financial-engineering-will-not-stabilise-unstable-euro-area.

De Marco, Filippo, and Marco Macchiavelli. 2016. "The Political Origin of Home Bias: The Case of Europe." FEDS Working Paper 2016-060.

European Commission. 2015. "Making the Best Use of the Flexibility within the Existing Rules of the Stability and Growth Pact." https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015DC0012&from=en.

European Fiscal Board. 2019. *Assessment of the European Fiscal Rules.* Brussels: European Fiscal Board.

Deslandes, J., and M. Magnus. 2019. "Which Supervisory or Regulatory Treatment of Banks' Exposures to Sovereign Risks?" https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/624434/IPOL_BRI(2019)624434_EN.pdf

Farhi, Emmanuel, and Iván Werning. 2017. "Fiscal Unions." *American Economic Review* 107 (12): 3788–3834

Feld, Lars, Christoph Schmidt, Isabel Schnabel, and Volker Wieland. 2018. "Refocusing the European Fiscal Framework." *VoxEu*, September 12. https://voxeu.org/article/refocusing-european-fiscal-framework.

Galí, Jordi, and Tommaso Monacelli. 2008. "Optimal Monetary and Fiscal Policy in a Currency Union." *Journal of International Economics* 76 (1): 116–32.

Garicano, Luis. 2019. "Two Proposals to Resurrect the Banking Union: The Safe Portfolio Approach and SRB+." https://www.bruegel.org/wp-content/uploads/2020/02/Two-proposals-to-resurrect-the-Banking-Union-by-Luis-Garicano.pdf.

German Council of Economic Experts. 2015. *The Long Road to More Financial Stability in Germany and in Europe.* Wiesbaden, Germany: German Council of Economic Experts.

Gourinchas, Pierre-Olivier., Philippe Martin, and Todd E. Messer. 2020. "The Economics of Sovereign Debt, Bailouts and the Eurozone Crisis." NBER Working Paper 27403

Horvath, Bálint L., Harry Huizinga, and Vasso Ioannidou. 2015. "Determinants and Valuation Effects of the Home Bias in European Banks' Sovereign Debt Portfolios." CEPR Discussion Paper 10661.

Italian Ministry of the Economy. 2015. "Documento di Economia e Finanza, Sezione I." http://www.dt.mef.gov.it/modules/documenti_it/analisi_programmazione/documenti_programmatici/SEZIONE_I_-_Programma_di_Stabilitx_xdeliberatox_on-line.pdf.

James, Harold. 2012. *Making the European Monetary Union.* Cambdridge, MA: Harvard University Press.

Kim, Jun Il, and Jonathan D. Ostry. Forthcoming. "Boosting Fiscal Space: The Roles of GDP-linked Debt and Longer Maturities." *Economic Policy.*

Martinez, Joseba, Thomas Philippon, and Markus Sihvonen. 2019. "Does a Currency Union Need a Capital Market Union? Risk Sharing via Banks and Markets." NBER Working Paper 26026.

Mohl, Philipp, and Gilles Mourrel. 2020. "Performance of the EU spending Rule: A Quantitative Assessment." Paper presented at the joint conference of European Fiscal Board, Amsterdam Centre for Economic Studies and CEPR on Rethinking the European Fiscal Framework, Brussels, February 28..

Panizza, Ugo. 2013a. "Sovereign Defaults and the Rules of Engineering." *VoxEu*, March 2. https://voxeu.org/article/sovereign-default-and-rules-engineering.

Panizza, Ugo. 2013b. "Incentives for Avoiding Delayed Sovereign Defaults." *VoxEu*, March 3. https://voxeu.org/article/incentives-avoiding-delayed-sovereign-defaults.

Porter, Eduardo. 2021. "How the American Unemployment System Failed." *New York Times*, January 21. https://www.nytimes.com/2021/01/21/business/economy/unemployment-insurance.html.

Schäuble, W. 2017. "Non-paper for paving the way towards a Stability Union." http://media2.corriere.it/corriere/pdf/2017/non-paper.pdf.

Scholz, O. 2019. "Position Paper on the Goals of the Banking Union." http://prod-upp-image-read.ft.com/b750c7e4-ffba-11e9-b7bc-f3fa4e77dd47.

Shiller, Robert J. 1993. *Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks.* Oxford: Oxford University Press.

Shiller, Robert, Jonathan D. Ostry, James Benford, and Mark Joy. 2018. "Sovereign GDP-linked bonds: Rationale and Design." *VoxEu,* March 16. https://voxeu.org/article/sovereign-gdp-linked-bonds-rationale-and-design.

Sunstein, Cass R. 1995. "Problems with Rules." *California Law Review* 83 (4): 952–1026.

Tabellini, Guido. 2017. "Reforming the Eurozone: Structuring versus Restructuring Sovereign Debts." *VoxEu,* November 23. https://voxeu.org/article/structuring-versus-restructuring-sovereign-debts-eurozone.

Tabellini, Guido. 2018. "Risk Sharing and Market Discipline: Finding the Right Mix." *VoxEu*, July 16. https://voxeu.org/article/risk-sharing-and-market-discipline-finding-right-mix.

Véron, Nicolas. 2017. *Sovereign Concentration Charges: A New Regime for Banks' Sovereign Exposures.* Bruegel & Peterson Institute for International Economics.

Wyplosz, Charles. 2013. "Fiscal Rules: Theoretical Issues and Historical Experiences." In *Fiscal Policy after the Financial Crisis*, edited by Alesina, Alberto and Francesco Giavazzi, 495–525. Chicago: University of Chicago Press.

An Ounce of Prevention

Joseph P. Newhouse

Benjamin Franklin (1735) advised that the marginal rate of substitution between prevention and cure had a constant value of –16, although he didn't put it in exactly those terms. Perhaps with Franklin's adage in mind, when discussions turn to the cost of health care, many advocate more resources for prevention.

In this paper, I look at prevention through an economic lens and make three main points. First, those advocating preventive measures are often asked how much money a given measure saves. This question is misguided. Instead, preventive measures can be thought of as insurance, with a certain cost in the present that may or may not pay off in the future. Although most medical preventive measures improve expected health, they do not save money. Various lifestyle and early childhood interventions, however, may both save money and improve health.

Second, preventive measures, including medical and lifestyle measures, are heterogeneous in their value, both across measures and within measure, across individuals. As a result, generalizations in everyday discourse about the value of prevention can be overly broad.

Third, health insurance coverage for medical preventive measures generally should be more extensive than coverage for the treatment of a medical condition, though full coverage of preventive services is not necessarily optimal.

■ *Joseph P. Newhouse is John D. MacArthur Professor of Health Policy and Management, Harvard University, Cambridge, Massachusetts and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His e-mail address is joseph_newhouse@ harvard.edu.*

The COVID-19 pandemic has made prevention highly salient; virtually the entire world has engaged in various forms of preventive behavior, including partial shutdowns of the economy, sheltering in place, wearing face masks, and frequent handwashing. Such behavior is not surprising when there is no immunity against a pathogen that is rather easily transmissible and potentially fatal.

Here, however, I focus on preventive measures against chronic diseases rather than infectious diseases like COVID-19. Chronic diseases account for the great bulk of American health care spending; pre-pandemic, persons with one or more chronic diseases accounted for 90 percent of American health care spending, although they were only 60 percent of the population (Buttorff, Ruder, and Bauman 2017). The pandemic has surely increased the proportion of health spending on infectious disease, but at some point it will recede and the chronic diseases will remain. Examples of preventive measures that I have in mind, in the case of coronary heart disease, include having one's blood pressure and cholesterol checked regularly and starting treatment if abnormal, maintaining a healthy diet, getting regular exercise, and keeping one's weight at reasonable levels.

An Economic Formulation of Preventive Measures

I define prevention as a medical or behavioral action that on expectation reduces either the probability or the potential severity of ill health in the non-immediate future. (The qualification of "non-immediate" is meant to exclude medical treatments that simply alleviate a symptom or other condition in the immediate future, such as taking aspirin for a headache.) This definition includes what epidemiologists call primary prevention, meaning an activity that reduces the likelihood of the initial occurrence of a disease, like a flu shot (or other vaccinations) or avoiding crowds during a pandemic. It also includes what epidemiologists term secondary prevention or reducing the likelihood of recurrence or the future severity of a disease one already has: for example, taking a drug that lowers blood pressure to reduce the likelihood of a second heart attack, as well as screening programs such as mammography that are intended to detect disease at an early stage when treatment may be more successful. Moreover, this definition includes both medical interventions as well as what Ehrlich and Becker (1972) term self-protection, or, in this context, maintaining a healthy lifestyle.

Despite the widespread belief among the general public that medical preventive measures save money, they frequently do not. Even more important, whether a preventive measure is welfare-increasing does not depend on whether it saves money any more than the welfare judgement on whether or how to treat a disease such as cancer depends on whether it saves money. Both medical prevention and medical treatment sometime save money, but often neither do.

From an economic perspective, a preventive activity is effectively an insurance policy that requires some cost in the present for a future benefit with an expected

positive value.[1] As a basic analytical framework, imagine a utility function that depends on consumption and health over a number of time periods, with a discount factor applied in future time periods. This utility function is optimized subject to the present value of a lifetime income constraint. To keep matters simple, it is useful as a starting point to assume perfect capital markets, and to ignore time constraints, uncertainty about one's lifetime, and bequest motives. Because of several uncertainties described below, an expectation must be formed for the present value of this utility function.

Next, add to this framework a preventive action that may be taken in the present, which is costly in terms of money, time, clinical risk, and/or disutility such as pain or physical discomfort. Then, the question is whether the expected present value of this consumption-and-health utility function is higher when taking the preventive action in the present or not taking it.[2]

In this framework, the formation of expectations is meant to capture several dimensions of uncertainty about the preventive action. First, the expected value of a preventive action will generally depend upon the future state of the world: For example, is the type of prevalent flu in the flu season one that those responsible for formulating the annual flu vaccine guessed will occur? Second, even if the preventive measure lowers risk as intended, the event in question may still occur. For example, persons who successfully lower their blood pressure and reduce their risk of a heart attack may still have a heart attack. Third, with some probability, the preventive measure itself may have adverse effects, such as a reaction to a vaccine.

Whereas the intent of conventional insurance is to protect against financial risk or variation in income level, the intent of a preventive measure is to protect against health risk or variation in health level. Furthermore, because health and earnings capability are related, if a preventive measure reduces future sick time and/or raises future productivity, it could increase lifetime income and the future stream of consumption, just as the smoothing of consumption from conventional insurance could affect health states. The analogy from conventional insurance to preventive health care is not exact, however, because with conventional insurance, the most one can lose is the premium paid, while with preventive medicine, there is also the possibility that the costs could include an adverse health reaction.

This framework for thinking about preventive health care measures can account for competing risks. For example, preventing a cardiac event may allow an individual to survive but then later to develop dementia and incur related costs. If the discounted expected utility of living with dementia is deemed sufficiently low, the preventive measure might conceivably decrease welfare.

The economic perspective of prevention as insurance that may increase or decrease welfare differs from the perspective of many medical and public health

---

[1] For a more extensive discussion of many of the points made in this paper, see Phelps (1978) and Kenkel (2000).

[2] This setup is a simplified statement of Michael Grossman's (1972, 2000) classic model of the demand for health.

professionals. A primary goal of medical professionals is to alleviate suffering and the consequences of disease, often expressed without regard to resource constraints. For example, a modern version of the Hippocratic Oath, as formulated in 1964, says in part: "I will prevent disease whenever I can, for prevention is preferable to cure." The version of the Oath that contains this sentence is the most popular form of the Oath among American medical school graduates, with one-third of the graduates taking it (Hajar 2017).

Heterogeneity across Individuals

Decisions about whether a preventive measure increases utility will differ across individuals for various reasons including degree of risk aversion, age, and other individual background factors that can predict health outcomes.

For example, in making decisions about preventive medicine, the degree of risk aversion is relevant. It is quite possible that the expected net benefit of taking a preventive action could be negative for a risk-neutral individual but positive for a risk-averse individual. Individuals may also vary in their choice of discount rates.

Decisions about preventive measures, of course, are taken throughout an individual's lifetime. Thus, given a finite lifetime, expected net benefits of preventive measures fall with age. It is not surprising, therefore, that guidelines for the use of medical preventive services account for age. For example, the US Preventive Services Task Force (2018a) recommends against prostate cancer screening for men age 70 and over. And the US Preventive Services Task Force (2016) recommends biennial mammography for women only up to age 74, concluding that "the current evidence is insufficient to assess the balance of benefits and harms of screening mammography in women 75 and older."

Heterogeneity across individuals also matters in creating clinical guidelines. Those responsible for formulating them prefer to base the guidelines on clinical trial results, but heterogeneity across individuals may cause results in practice to differ from those of a clinical trial. Einav et al. (2019) point out that the population in a trial of a preventive measure, for example all females of a given age range, may differ in its risk distribution from those who choose to obtain a preventive measure such as mammography. For example, those who choose to obtain a mammography may have knowledge that they are personally more susceptible to breast cancer because of habits like smoking or background like family genetic history.

Heterogeneity among Clinical Measures of Effectiveness of Preventive Medicine

The utility-maximizing framework discussed here follows most of the American health economics literature in using standard welfare economics, which can also be translated into cost-benefit analysis comparing the costs of preventive medicine to

expectations of possible gains in consumption and health. Unfortunately, however, the empirical health economics literature has only scattered results that quantify the benefits of specific preventive measures. There is, however, an extensive public health literature that uses Quality Adjusted Life Years (QALYs) or Disability Adjusted Life Years (DALYs) to assess the value of various clinical preventive measures (Torrance 1986; Gold et al. 1996; Neumann et al. 2016). QALYs and DALYs are utility-weighted life expectancies that account for an individual's state of health or quality of life at each point in time, rather than simply expected length of life using stated-preference methods to elicit weights. A year of perfect health is weighted as one, death is weighted as zero, and a year spent with a chronic illness—including a range of conditions like hay fever, cancer in remission, or a specified level of disability—is given an intermediate value. In QALY terms, therefore, the usual life-expectancy measure assumes all remaining years are spent in perfect health. The cost-effectiveness of a preventive or treatment measure is its incremental cost per incremental QALY, relative to the next-best measure.

If the preventive measure does not save money and is financed from an insurance pool or is tax-financed, one faces the question of whether its health benefits are worth the cost. Here, the public health literature departs from standard welfare economics, which is based on preferences as revealed by an observed demand curve; that is, in standard welfare economics, an individual compares personally paying the costs of preventive care (broadly understood to include both monetary and non-monetary costs) with the expected personal benefits received in terms of future consumption and health. This approach will clearly depend on an individual's willingness to pay, which will vary with income.

In contrast, the public health literature typically gives QALYs and DALYs equal weight across persons and thus removes willingness to pay from the calculation of incremental cost per incremental QALY. In principle, one could weight persons according to a measure of social welfare (Saez and Stantcheva 2016), but this is usually not done in practice. The public health literature often rationalizes the use of equal weights on equity grounds, but one can also appeal to the extensive literature that questions the normative meaning of observed demand curves, much of which uses health care choices under uncertainty as examples (Handel and Kolstad 2015; Bhargava, Loewenstein, and Sydnor 2017; Ericson and Sydnor 2017; Handel and Schwartzstein 2018).

Having computed the incremental cost per incremental QALY for a given measure, the public health literature evaluates whether the measure should be implemented by determining if its incremental cost per incremental QALY falls below a specified dollar threshold. A common rule-of-thumb threshold in US studies is $100,000 per QALY, although this may well now be a lower bound; in 2020, a value of $150,000 or even $200,000 may be more realistic (Braithwaite et al. 2008).

The key points for my purposes, however, are the large variation in cost per QALY across various preventive measures found in the public health literature and the modest number of measures—about 20 percent—that both improve health and reduce cost (Cohen, Neumann, and Weinstein 2008). Vaccination is a well-known

Table 1

Examples of Cost-Effectiveness Ratios of Selected Preventive Measures
(in 2006 dollars)

Cost-Saving	Polypill versus individual prescriptions and their rate of usage of up to four medications from recommended medication classes for secondary prevention of cardio-vascular disease
Cost < $100,000/QALY	Newborn screening for medium-chain acyl-coenzyme A dehydrogenase deficiency, $160/QALY
	High intensity smoking-relapse program, $190/QALY
	Intensive tobacco use prevention program for 7th and 8th graders, $23,000/QALY
Cost between $100,000–$1,000,000/QALY	Screening all 65-year olds with hypertension for diabetes, $590,000/QALY
Cost > $1,000,000/QALY	Antibiotic prophylaxis (amoxicillin) for children with moderate cardiac lesions who are undergoing urinary catheterization, $10,000,000/QALY

Source: These examples were drawn on June 15, 2020 from the 100 publicly accessible values of the registry of cost-effectiveness studies maintained by the Center for the Evaluation of Value and Risk in Health at Tufts Medical Center, http://healtheconomics.tuftsmedicalcenter.org/cear2n/search/ratio0.aspx. As of June 2020, the entire Tufts registry contained results of over 20,000 cost-effectiveness studies, but the public website only shows the 100 that have been most recently added.

example of a measure that improves health and reduces cost. It is typically inexpensive, causes few adverse events, and can confer immunity for many years. The development of the polio vaccine, for example, was one of the great public health triumphs of the 20th century. In the late 1940s, polio crippled 35,000 Americans annually; because of vaccination, it was eradicated in the United States in 1979 (Centers for Disease Control and Prevention 2020). Vaccination also differs from many other preventive measures because of the external benefit it confers on the unvaccinated ("herd immunity"). Another example of a preventive measure that saves money and improves health is a "polypill"—a single pill with several active ingredients for secondary prevention of heart disease versus single prescriptions for various agents (Gaziano et al. 2019).

The remaining 80 percent of preventive measures do not save money. Table 1 shows a few examples of preventive measures. The majority of all preventive measures—about 60 percent of them—provide health benefits at a cost of less than $100,000/QALY (2006 dollars). Another 10 percent of measures cost between $100,000 and $1,000,000 per QALY; those measures with costs near the lower end of this range might pass the common rules of thumb of cost-effectiveness described above. The remaining 10 percent of preventive measures studied in the literature either worsen expected health or, if they improve it, cost more than $1,000,000 per QALY.

An important caveat is that the published ratios for cost-per-QALY shown in Table 1 are average rather than marginal or incremental values. However, the use of average values for policy purposes is mitigated to some degree because many of the measures are both studied and described as being conditioned on observables such as age, sex, and prior disease history, and within those subgroups, the expected benefit may be approximately constant.

Errors in Screening Procedures

Many common preventive procedures, like mammography, screen a population for evidence of disease. Errors in such measures can be large enough to make them welfare-decreasing, at least among certain subgroups. Errors can be both false positives (known as "Type 1" errors), in which a screening test may signal disease when it is not present, or false negatives (known as "Type 2" errors), in which a screening test may fail to detect disease when it is present.

For example, in one large observational study, mammography gave an abnormal result 11.6 percent of the time but had a false positive rate of 95.6 percent (2007–2013 data) (Lehman et al. 2017).[3] False positive rates depend heavily on the true prevalence of the disease in the sample studied. Other studies with different populations give lower but still relatively high false positive rates for mammography (Nelson et al. 2016). Using the 95.6 percent figure, however, suggests that for every 1,000 women screened, mammography would signal that further investigation was indicated for 116 of them, but 111 of those 116 women would not actually have breast cancer. The likelihood of a false positive together with remaining life expectancy is one reason that US Preventive Services Task Force (2016) concluded that evidence was insufficient to recommend screening mammography for women 75 and older.

A screening test does not generally diagnose the illness but merely suggests its presence. If the screening test is positive, the next step is either to repeat the test or carry out a diagnostic test or procedure. Repeating the test is more attractive the higher the false positive rate, the lower the cost of the test, and the greater the independence between the results of the two tests. In the mammography example, a

---

[3] Of 1,682,504 mammograms in the Lehman et al. (2017) study, 194,668 had an interpretation of an abnormality (0.116 = 194,668/1,682,504). Among the 194,668 women who screened positive, there were 8,529 true positives and 186,140 false positives. The false positive rate was therefore 0.956 = 186,140/194,668. Among the 1,487,836 women who screened negative, there were 1,283 false negatives and 1,486,663 true negatives. In epidemiologic jargon, the "sensitivity" of mammography (equal to the ratio of screen positive/true positive), was 86.9 percent and the "specificity" (equal to the screen negative/true negative) was 88.9 percent. The "true prevalence" in the population (equal to the true positive rate plus the false negative rate) was (8,529 + 1,283)/1,682,504 = 0.0058 percent. The calculation that 111 of the 116 women who screened positive for breast cancer did not actually have it is 111 = (186,140/194,668) × 116. The false negative rate was 1,283/1,487,836.

common follow-up diagnostic procedure is a biopsy, which of course has non-trivial expense and can result in medical complications.

Continuing the results from the large observational study of mammography (Lehman et al. 2017), 884 of the 1,000 women screened would have had a negative test. In the study, mammography had a false negative rate of 0.09 percent, so using expected values, one of those 884 women would actually have breast cancer. Both because of the possibility of false negatives and even more because of the probability of a woman's developing detectable breast cancer in a given time interval after a negative screen, periodic testing is desirable. More frequent testing, however, increases the number of women experiencing false positives. Hubbard et al. (2011) and Nelson et al. (2016) calculate that among women who hypothetically would have had annual screening mammography for a decade, 61 percent would have had at least one false positive screening and that biennial (every two years) testing would lower this rate to 42 percent. As noted above, the US Preventive Services Task Force (2016) recommends biennial testing for women 50–74.

In sum, the optimal screening frequency depends on several parameters, especially the true prevalence in the population, the frequency and cost of Type 1 and Type 2 errors, the cost of the test itself, and the probability and value of successful prevention or treatment if the screening test is positive. A sufficiently high false positive rate can make the value of a screening test negative, depending on the cost of the test, the cost and potential adverse health consequences of any follow-up procedures from a positive test, and whether treating the disease can be just as successful if the disease is detected at a later time when a person has symptoms.

Insurance Coverage of Preventive Services

At one time, it was common to hear arguments that clinical preventive services were not insurable because they were "not a random variable and hence not an 'insurable risk'" (Zweifel and Breyer 1997). Zweifel and Breyer give the example that "it is hardly conceivable that a health insurer would ever cover expenditure on items such as . . . atomizers that help to prevent respiratory disorders;" a similar point could be made about a flu shot or mammography. There are, however, both economic efficiency and behavioral arguments for many preventive measures.

Standard Efficiency Arguments

A textbook example of the efficiency argument for clinical preventive services is the externality from vaccination against a contagious disease mentioned above. The externality is the rationale for compulsory measles vaccination (Oster 2018). In part, because of the externality, billions of dollars are being invested in 2020 to develop a vaccine for COVID-19.

Even if the contagion argument is not relevant, however, there are standard efficiency arguments for not only insuring preventive services but subsidizing them to a greater degree than treatment services. Assume that certain kinds of

prevention do in fact reduce the likelihood of disease and thus the demand for treatment; in other words, preventive services substitute for treatment services. Also assume that an individual's insurance premiums or taxes are negligibly affected by their personal use of preventive and treatment services, as is generally the case. In this setting, preventive services should be at least partially insured. Otherwise individuals, by not accounting for the financial consequences of the reduction in their future use of treatment services, will under consume preventive services relative to a social optimum (Ellis and Manning 2007; Goldman and Philipson 2007).

Although this argument does not imply that full insurance is optimal, the 2010 Patient Protection and Affordable Care Act made all preventive services rated A and B by the US Preventive Services Task Force free for public insurance plans as well as the great majority of private insurance plans (the exceptions are so-called grandfathered private plans, but few of those remain). An A rating means the service has a high certainty of substantial net benefit, while a B rating means it has a moderate certainty of substantial net benefit or high certainty of moderate to substantial net benefit (US Preventive Services Task Force 2018b).

The logic of less cost-sharing for preventive services also implies that drugs to treat a chronic disease such as diabetes should face less cost-sharing than drugs to reduce a transient problem such as itching or pain from a sprain or minor burn, because failure to take the diabetic drug could lead to exacerbations of the disease that would be costly to treat, whereas a transient problem would resolve with time and not require further treatment. In practice, however, insurance policies generally do not differentiate cost-sharing for drugs according to the potential consequences of noncompliance. This has led to proposals to make medications that treat serious chronic diseases such as diabetes and cardiac conditions free to the consumer to forestall downstream adverse events, under a philosophy of "value-based insurance design" (Chernew et al. 2008). Some steps in this direction have been taken by Medicare Advantage plans (Starc and Town 2018).

Behavioral Considerations

Behavioral considerations offer another—and likely more powerful—rationale for insurance coverage of preventive activities. Failure to engage in preventive activities is a standard example in the hyperbolic discounting literature. Every day, I may eat a cookie and plan to exercise tomorrow, rather than the other way around. In this situation, decisions are not time-consistent, and people will later experience regret that they did not take certain actions at earlier times.

It is clear that time inconsistency can be an issue in health-related decisions, like smoking. In a classic paper in this genre, Gruber and Koszegi (2001) begin from the rational addiction model of Becker and Murphy (1988), in which tobacco use is individually rational if individual decisions are time consistent. Gruber and Koszegi show that if individual decisions are not time-consistent, one can make a case for much higher excise taxes than taxes that simply equal the external costs because those high taxes will reduce the "internalities" experienced from smoking

by a time-inconsistent individual. The evidence for time-inconsistent behavior is not conclusive, but in the case of smoking, some support for the hypothesis can be found in the regret expressed by many addicted smokers—many of whom initiated tobacco use in adolescence before the legal age for smoking—and the corresponding demand for aids to quit smoking.

However, the evidence on whether more extensive insurance coverage overcomes the issue of time inconsistency by reducing the up-front cost for preventive measures suggests that reducing out-of-pocket cost is not the main barrier. In one randomized trial, insured patients who had suffered a heart attack received free access to four types of drugs to reduce the likelihood of a second and possibly fatal heart attack. Meanwhile, the control group continued on their existing insurance plan with their usual copayment for drugs (Choudhry et al. 2011). Rates of adherence increased with free drugs but only by about 5 percentage points. Even more strikingly, over one-half of the population in both the treatment and control groups did not comply with prescriptions for the drugs, despite the potentially dire consequences of noncompliance. These findings echoed those from the RAND Health Insurance Experiment; in that study, when both preventive and treatment services were free, rates of compliance with preventive guidelines were around 10 percentage points higher than when services were costly to the patient, but compliance was still well under half for adult males, although it was higher for females (Newhouse and the Insurance Experiment Group 1993).

Behavioral considerations may shape the structure of dental insurance, the most common form of which is a group policy with relatively low annual benefit limits. Dental insurance almost always, however, covers prophylaxis (teeth cleaning). Thus, rather than offering much protection against expensive dental procedures, dental insurance can be seen as a nudge to seek preventive care. One can only speculate as to why dental insurance has such low limits, but one possibility is that because dental insurance was and remains mainly provided by dental societies (Delta Dental), and dentists may well have wanted a mechanism to encourage patients to seek care through an untaxed employer-provided fringe benefit, while still having the freedom to price expensive dental procedures.[4] Although the Patient Protection and Affordable Care Act of 2010 banned annual and lifetime upper limits on payouts from medical insurance, it did not apply those limits to dental insurance.

---

[4]Data from the RAND Health Insurance Experiment (Newhouse and the Insurance Experiment Group 1993) show that the probability of using diagnostic and preventive dental services in a year is 15 percentage points higher when care is free than when it is costly. My own view is that the current situation is a low-level equilibrium trap; because dental insurance is mostly employment-based, any individual employer that offered more generous dental insurance benefits (for example, for orthodontia), could well be selected against in the labor market. Cabral (2017) has shown that individuals can time their dental expenditures, which potentially creates a selection problem in individual markets, but timing should not be much of an issue for self-insured employers with relatively low employee turnover (such an employer doesn't much care whether the dental work is done in December or January), and many such employers offer dental insurance to employees and their dependents.

Table 2

American Death Rates from the Ten Leading Causes of Death in 1900 and 2016

(deaths per 10,000 population)

1900 Causes of Death/10,000		2016 Causes of Death/10,000	
Influenza and pneumonia	202.2	Diseases of the heart	16.6
Tuberculosis	194.4	Malignant Neoplasms (cancer)	15.6
Diarrhea, Enteritis, and ulceration of the intestines	142.7	Unintentional injuries	4.7
Diseases of the heart	137.4	Chronic lower respiratory diseases	4.1
Intracranial lesions of vascular origin	106.9	Cerebrovascular diseases	3.8
Nephritis	88.6	Alzheimer's disease	3.0
All accidents	72.3	Diabetes mellitus	2.1
Cancer and other malignant tumors	64.0	Suicide	1.4
Senility	50.2	Nephritis, nephrotic syndrome, and nephrosis	1.3
Diphtheria	40.3	Influenza and pneumonia	1.1

Source: https://www.cdc.gov/nchs/data/dvs/lead1900_98.pdf and https://www.cdc.gov/nchs/hus/contents2017.htm#Table, Table 17. The 1900 data are from eleven death-reporting states, mostly in the Northeast, plus the District of Columbia. The specific states can be found in Table 1.04 in Appendix II of https://www.cdc.gov/nchs/data/misc/usvss.pdf.

Non-Clinical Preventive Interventions

Over the course of the 20th century, there was a remarkable shift in the leading causes of death from acute to chronic diseases, as illustrated in Table 2 by the ten leading causes of death in the United States in 1900 and 2016. In 1900, the three leading causes of death were attributable to infectious disease, and two of those three are not even among the top ten causes in recent years. Influenza and pneumonia, the leading cause of death in 1900, was the tenth leading cause in 2016, and death rates from it have fallen by a factor of 200. (Deaths from pneumonia will be higher in 2020 because of COVID-19, but once the pandemic recedes, the causes of death should again be similar to those shown in Table 2.) Many of the chronic diseases that are among the current leading causes of death are influenced by lifestyle and health habits, including tobacco use, diet, exercise, and substance misuse. While rates of smoking have fallen, the ongoing, worsening obesity epidemic suggests that policy efforts to improve diet and exercise have had at best modest effects (Cutler, Glaeser, and Shapiro 2003; Ward et al. 2019).

Clinical preventive care plays a role in addressing these health issues, perhaps especially in the case of screening for cancer, high blood pressure, high cholesterol, and depression. Some of the most significant preventive interventions to improve health status in the 20th and 21st century, however, have been nonmedical.

Taxing or Regulating Goods with High Health Costs

The flip side of subsidizing preventive medical services is taxing unhealthy goods or services to reduce their use. Public policies to alter lifestyles and health habits have a mixed track record. The most successful are likely the increased cigarette taxes and laws and regulations prohibiting smoking in public places that have played a central role in the percentage of cigarette smokers falling from 42 percent in 1965 to 16 percent in 2016 among Americans age 18 and over (Chaloupka and Warner 2000; National Center for Health Statistics 2018). That large decline in smoking has played a major role in the fall of death rates from lung cancer among males—roughly a factor of two since 1990. Lung cancer rates among females have also begun to decline but to a lesser extent because female smoking rates declined later than rates among males (Siegel, Miller, and Jemal 2020).

Externalities are a standard rationale for tobacco excise taxes. Their value can be quantified by assuming two cohorts that differ only in that one begins to smoke at age 20, while the other does not, and then calculating the present value of external costs in the two cohorts. Assume both cohorts pay a common payroll tax to finance both health insurance and a pension. On one side, tobacco users impose costs on non-users because taxes must be higher to finance their additional health costs. On the other side, tobacco users subsidize non-users in a pension system because their lower life expectancy means that they collect less in pension benefits when or if they become eligible. The undiscounted difference in pension benefits between the two cohorts is substantial, but because the pension benefits are zero for the first 40+ years after smoking is assumed to begin, they are heavily discounted with a typical discount rate. A lower discount rate will make the long-term pension effects look larger, and the reverse is also true. Using 1980s American data, the two effects were equal at a real discount rate of a little over 3 percent (Manning et al. 1989). Coincidentally, the 3 percent real rate is the rate the two Panels on Cost-Effectiveness in Health and Medicine have recommended be used in cost-effectiveness studies (Gold et al. 1996; Neumann et al. 2016). As mentioned above, however, if individuals are time-inconsistent, the optimal tax is much higher because of the size of the "internality."

Analogous to tobacco, sugar-sweetened beverages have been linked to obesity, cardiovascular disease, and diabetes, which increase health care costs, and some localities have begun to tax them. In this journal, Alcott, Lockwood, and Taubinsky (2019) thoroughly discuss the issues around taxation of sugar-sweetened beverages.

Alcohol also imposes external and internal costs, but the case for a tax to improve economic efficiency is more complicated than for tobacco because alcohol in moderation may have beneficial health effects. As a result, the magnitude of an optimal efficient tax on alcohol is uncertain (Manning et al. 1989; Pogue and Sgontz 1989). In addition to the external costs involved in smoking, alcohol has large external costs both from driving while intoxicated and from its positive effect on violent crime (Chalfin, Hansen, and Ryley 2019).

Exercise and Workplace Wellness Programs

Workplace wellness programs are a form of employment-based subsidy to preventive activities that seek to promote healthy lifestyles among a firm's workers. Employers offering such programs may be motivated by selection concerns with respect to their labor force, a desire to increase the productivity of their labor force, workers' desire for a non-taxable fringe benefit, or some combination. Song and Baicker (2019) implemented a randomized trial of such a program at a large retail warehouse company. The trial consisted of eight modules implemented over 18 months emphasizing nutrition, physical activity, stress reduction, and prevention. Like subsidizing preventive care through health insurance, results suggested little effect on behavior. In the intervention group, somewhat more persons reported engaging in regular exercise (8 percentage points) and actively managing their weight (14 percentage points), but a large number of other pre-specified outcome measures did not differ significantly between the intervention and control groups. These included 27 self-reported measures of health outcomes and behaviors, including sleep quality and food choices; ten clinical markers, including cholesterol, blood pressure, and body mass index; 38 medical and pharmaceutical spending and utilization measures; and three employment outcomes, including absenteeism, job performance, and job tenure. Jones, Molitor, and Reif (2019) implemented a similar randomized trial among employees of the University of Illinois at Urbana-Champaign and, like Song and Baicker, found essentially null results.

Childhood Interventions

Many chronic diseases can be traced back to childhood deprivation. Childhood interventions with disadvantaged children, especially early childhood interventions, appear to have important effects not only on child health but also on health as an adult—especially for males (Heckman, Pinto, and Savelyev 2013; Campbell et al. 2014; Hendren and Sprung-Keyser 2019; García et al. 2020). For example, the Carolina Abcedarian Project involving children born between 1972 and 1977 provided cognitive and social stimulation throughout an eight-hour day as well as health care to a small randomized group of preschool, disadvantaged children age 0–5, and compared them with a control group. At a 30-year follow-up, the researchers obtained data on 19 of the original 29 treatment group males and nine of the original 23 control group males. Those with missing data, however, appeared to be missing at random. Despite the small numbers, several risk factors were markedly lower in the treatment group males around 30 years later, including obesity and hypertension, dyslipidemia (high cholesterol), and metabolic syndrome (a combination of obesity, cardiovascular disease, diabetes).

In a broader study, Goodman-Bacon (2018) has shown that the introduction of Medicaid a half-century ago reduced infant and child mortality. Hendren and Sprung-Keyser (2019), drawing on the work of Wherry, find positive effects of the more recent Medicaid expansions to cover older children; indeed, the effects from expanding Medicaid coverage in childhood are more than repaid by lower health care spending later in life. Hoynes, Schanzenbach, and Almond (2016) review a

considerable literature showing that malnutrition in utero or in early childhood leads to poorer adult health outcomes and go on to show that access to food stamps in utero or in early childhood reduces the prevalence of metabolic syndrome in adulthood. A thorough review of policies to reduce child poverty can be found in National Academies of Sciences, Engineering, and Medicine (2019).

The Supply of Preventive Services

The discussion up to this point has focused on demand for preventive measures or on justifications for their cost-effectiveness, but supply considerations are also relevant. For example, firms investing in research and development may tend to favor products to treat rather than to prevent disease.

One reason is that a firm will tend to prefer products that can be tested in a clinical trial with shorter durations because returns will come sooner; indeed, Budish, Roin, and Williams (2015) provide evidence that cancer clinical trials favor shorter-term projects. Moreover, such trials are more likely to be for agents designed to treat rather than to prevent various cancers because trials of treatment agents will typically need less time than those of prevention agents to establish whether an agent is efficacious. Even within trials of agents to treat cancer, the sample population will often be those with late-stage cancers because results will be available sooner. More generally, because trials of preventive agents are in healthy populations, they will likely take longer to show a response (or not) than a treatment for a sick person.

Kremer and Snyder (2015) illustrate how heterogeneity in a population's risk of contracting a disease also tends to favor trials of treatment agents rather than prevention agents. The intuition is simplest in the case of a monopolist deciding between two products that have the same research and development cost, show the same likelihood of success in development, are both perfectly effective, have no production cost or side effects, and are directed to a disease that results in the same harm for all who get the disease. The monopolist is assumed to know the distribution of risk in the population but not the risk of individual consumers. In the case of the treatment product, someone with the disease will be willing to pay up to the value of the harm the disease causes, while those who do not have the disease will not be willing to pay anything. In the case of the preventive product, consumers will pay up to their *expected* harm, but because of the varying risk among consumers, this amount will vary. The firm can choose to sell the preventive product only to higher-risk consumers at a price that equals their expected harm, in which case, lower-risk consumers will not buy, and the firm's revenue relative to the treatment case will be less. Alternatively, the firm can sell to all consumers at a price that equals the expected harm to low-risk consumers, but this generates less revenue than the treatment product because the firm sells the preventive product to the high-risk consumers at the price a low-risk consumer would pay. In the special case in which all consumers have the same risk of contracting the disease, the returns to the firm from the two types of products are equal.

Conclusion

Preventive activities are best viewed in conceptual terms as insurance where the payoff is a reduction in the likelihood of states of the world with lower utility. There are a vast array of preventive activities, with some purchased through the market, especially clinical interventions, some largely produced by the individual's time, such as exercise, and some publicly provided or subsidized. There is heterogeneity in value both across preventive measures as well as across individuals for a single measure.

The decision to bear the costs of preventive activity, like the decision to purchase conventional insurance, will not always pay off. In the case of conventional insurance, one may end up paying far more in premiums than one receives in claims. In the case of preventive decisions, one may pay the costs of clinical preventive care or shifts in personal habits, but still end up needing to be treated for the medical condition.

In public rhetoric, a common test applied to preventive activity is to ask whether it saves money, usually from avoided medical treatment, but this test is much too narrow because it omits potential future benefits of improved health and productivity. Only a minority of clinical preventive measures can be expected to save money in the narrow sense of reducing future health care spending, but many are worth their cost when expected gains from health and productivity are taken into account.

There are both standard efficiency and behavioral arguments for subsidizing preventive activities to a greater degree than treatment activities. The gist of the efficiency argument is that an individual ignores the cost imposed on others in the health insurance pool from not using preventive activities; the gist of the behavioral argument is hyperbolic discounting. But from a broad social perspective, some of the most important and high-payoff preventive activities may be outside clinical medicine, including policies such as cigarette taxes and smoking regulations as well as interventions to improve health and cognitive skills in early childhood, especially among disadvantaged children.

References

Alcott, Hunt, Benjamin B. Lockwood, and Dmitry Taubinsky. 2019. "Should We Tax Sugar-Sweetened Beverages? An Overview of Theory and Evidence." *Journal of Economic Perspectives* 33 (3): 202–27.

Becker, Gary S., and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy* 96 (4): 675–700.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor. 2017. "Choose to Lose: Health Plan Choices from a Menu with Dominated Options." *Quarterly Journal of Economics* 132 (3): 1319–72.

Braithwaite, R. Scott, David O. Meltzer, Joseph T. King Jr., Douglas Leslie, and Mark S. Roberts. 2008. "What Does the Value of Modern Medicine Say About the $50,000 per Quality-Adjusted Life-Year Decision Rule?" *Medical Care* 46 (4): 349–56.

Budish, Eric, Benjamin N. Roin, and Heidi Williams. 2015. "Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials." *American Economic Review* 105 (7): 2044–85.

Buttorff, Christine, Teague Ruder, and Melissa Bauman. 2017. *Multiple Chronic Conditions in the United States.* Santa Monica, CA: Rand.

Cabral, Marika. 2017. "Claim Timing and *Ex Post* Adverse Selection." *The Review of Economic Studies.* 84 (1): 1–44.

Campbell, Frances, Gabriella Conti, James J. Heckman, Seong Hyeok Moon, Rodrigo Pinto, Elizabeth Pungello, and Yi Pan. 2014. "Early Childhood Investments Substantially Boost Adult Health." *Science* 343 (6178): 1478–85.

Centers for Disease Control and Prevention. 2020. USA.gov. https://www.cdc.gov/polio/what-is-polio/polio-us.html (accessed June 23, 2020).

Chalfin, Aaron, Benjamin Hansen, and Rachel Ryley. 2019. "The Minimum Legal Drinking Age and Crime Victimization." NBER Working Paper Working Paper 26051.

Chaloupka, Frank J., and Kenneth E. Warner. 2000. "The Economics of Smoking." In *Handbook of Health Economics*, Vol. 1b, edited by A. J. Culyer and J. P. Newhouse, 1539–1627. Amsterdam: Elsevier.

Chernew, Michael E., Mayur R. Shah, Arnold Wegh, Stephen N. Rosenberg, Iver A. Juster, Allison B. Rosen, Michael C. Solol, Kristina Yu-Isenberg, and A. Mark Fendrick. 2008. "Impact of Decreasing Copayments on Medication Adherence Within a Disease Management Environment." *Health Affairs* 27 (1): 103–112.

Choudhry, Niteesh K., Jerry Avorn, Robert J. Glynn, Elliott M. Antman, Sebastian Schneeweiss, Michelle Toscano, Lonnie Reisman et al. 2011. "Full Coverage for Preventive Medications after Myocardial Infarctions." *New England Journal of Medicine* 365 (22): 2088–97.

Cohen, Joshua T., Peter Neumann, and Milton C. Weinstein. 2008. "Does Preventive Care Save Money? Health Economics and the Presidential Candidates." *New England Journal of Medicine* 358 (7): 661–63.

Cutler, David M., Edward L. Glaeser, and Jesse M. Shapiro. 2003. "Why Have Americans Becoming More Obese?" *Journal of Economic Perspectives* 17 (3): 93–118.

Ehrlich, Isaac, and Gary S. Becker. 1972. "Market Insurance, Self-Insurance, and Self-Protection." *Journal of Political Economy* 80 (4): 623–48.

Einav, Liran, Amy Finkelstein, Tamar Oostrom, Abigail J. Ostriker, and Heidi L. Williams. 2019. "Screening and Selection: The Case of Mammograms." NBER Working Paper 26162.

Ellis, Randall P., and Willard G. Manning. 2007. "Optimal Health Insurance for Prevention and Treatment." *Journal of Health Economics* 26 (6): 1128–50.

Ericson, Keith Marzilli, and Justin Sydnor. 2017. "The Questionable Value of Having a Choice of Levels of Health Insurance Coverage." *Journal of Economic Perspectives* 31 (4): 51–72.

Franklin, Benjamin. 1735. "Letter to the Editor of the Philadelphia Gazette." http://www.ushistory.org/franklin/philadelphia/fire.htm (accessed September 19, 2019).

García, Jorge Luis, James J. Heckman, Duncan Ermini Leaf, and María José´ Prados. 2020. "Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program." *Journal of Political Economy* 128 (7): 2502–41.

Gaziano, Thomas A., Ankur Pandya, Stephen Sy, Thihago Veiga Jardim, Jenna M. Ogden, Anthony Rodgers, and Milton C. Weinstein. 2019. "Modeling the Cost Effectiveness and Budgetary Impact of Polypills for Secondary Prevention of Cardiovascular Disease in the United States." *American Heart Journal* 214: 77–87.

Gold, Marthe R., Joanna E. Siegel, Louise B. Russell, and Milton C. Weinstein. 1996. *Cost-Effectiveness in*

Health and Medicine. New York: Oxford University Press.

Goldman, Dana P., and Tomas J. Philipson. 2007. "Integrated Insurance Design in the Presence of Multiple Medical Technologies." *American Economic Review* 97 (2): 427–32.

Goodman-Bacon, Andrew. 2018. "Public Insurance and Mortality: Evidence from Medicaid Implementation." *Journal of Political Economy* 126 (1): 216–62.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80 (2): 223–55.

Grossman, Michael. 2000. "The Human Capital Model." In *Handbook of Health Economics*, Vol. 1a, edited by A. J. Culyer and J. P. Newhouse, 347–408. Amsterdam: Elsevier.

Gruber, Joanthan, and Botond Koszegi. 2001. "Is Addiction Rational? Theory and Evidence." *Quarterly Journal of Economics* 116 (4): 1261–1305.

Hajar, Rachel. 2017. "The Physician's Oath: Historical Perspectives." *Heart Views* 18 (4): 154–59.

Handel, Benjamin R., and Jonathan T. Kolstad. 2015. "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review* 105 (8): 2449–2500.

Handel, Benjamin R., and Joshua Schwartzstein. 2018. "Frictions or Mental Gaps: What's Behind the Information We (Don't) Use and When Do We Care?" *Journal of Economic Perspectives* 32 (1): 155–78.

Heckman, James J., Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103 (6): 2052–86.

Hendren, Nathaniel, and Ben Sprung-Keyser. 2019. "A Unified Welfare Analysis of Government Policies." NBER Working Paper 26144.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. 2016. "Long-Run Impacts of Childhood Access to the Safety Net." *American Economic Review* 106 (4): 903–34.

Hubbard, Rebecca A., Karla Kerlikowske, Chris I. Flowers, Bonnie C. Yankaskas, Weiwei Zhu, and Diana L. Miglioretti. 2011. "Cumulative Probability of Falso-Positive Recall or Biopsy Recommendation after 10 Years of Screening Mammography." *Annals of Internal Medicine* 155 (8): 481–92.

Jones, Damon, David Molitor, and Julian Reif. 2019. "What Do Workplace Wellness Programs Do? Evidence from the Illinois Workplace Wellness Study." *Quarterly Journal of Economics* 134 (4): 1747–91.

Kenkel, Donald S. 2000. "Prevention." In *Handbook of Health Economics*, Vol. 1b, edited by A. J. Culyer and J. P. Newhouse, 1676–1720. Amsterdam: Elsevier.

Kremer, Michael, and Christopher M. Snyder. 2015. "Preventives versus Treatments." *Quarterly Journal of Economics* 130 (3): 1167–1239.

Lehman, Constance D., Robert F. Arao, Brian L. Sprague, Janie M. Lee, Diana S.M. Buist, Karla Kerlikowske, Louise M. Henderson, Tracy Onega, Anna N.A. Tosteson, Garth H. Rauscher, and Diana L. Miglloretti. 2017. "National Performance Benchmarks for Modern Screening Digital Mammography: Update from the Breast Cancer Surveillance Consortium." *Radiology* 283 (1): 49–58.

Manning, Willard G., Emmett B. Keeler, Joseph P. Newhouse, Elizabeth M. Sloss, and Jeffrey Wasserman. 1989. "The Taxes of Sin: Do Smokers and Drinkers Pay Their Way?" *JAMA* 261 (11): 1604–09.

National Academies of Sciences, Engineering, and Medicine. 2019. *A Roadmap to Reducing Child Poverty.* Washington, DC: The National Academies.

National Center for Health Statistics. 2018. *Health, United States, 2017: With Special Feature on Mortality.* Hyattsville, MD: US Department of Health and Human Services.

Nelson, Heidi D., Amy Cantor, Linda Humphrey, Rochelle Fu, Miranda Pappas, Monica Daeges, and Jessica Griffin. 2016. *Screening for Breast Cancer: A Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation.* Rockville, MD: Agency for Healthcare Research and Quality.

Neumann, Peter J., Gillian D. Sanders, Louise B. Russell, Joanna E. Siegel, and Theodore G. Ganiats, eds. 2016. *Cost-Effectiveness in Health and Medicine.* New York: Oxford University Press.

Newhouse, Joseph P., and the Insurance Experiment Group. 1993. *Free for All? Lessons from the RAND Health Insurance Experiment.* Cambridge: Harvard University Press.

Oster, Emily. 2018. "Does Disease Cause Vaccination? Disease Outbreaks and Vaccination Response." *Journal of Health Economics* 57: 90–101.

Phelps, Charles E. 1978. "Illness Prevention and Medical Insurance." *Journal of Human Resources* 13: 183–207.

Pogue, Thomas F., and Larry G. Sgontz. 1989. "Taxing to Control Social Costs: The Case of Alcohol." *American Economic Review* 79 (1): 235–43.

Saez, Emmanuel, and Stefanie Stantcheva. 2016. "Generalized Social Welfare Weights for Optimal Tax Theory." *American Economic Review* 106 (1): 24–45.

Siegel, Rebecca L., Kimberly D. Miller, and Ahmedin Jemal. 2020. "Cancer Statistics, 2020." *CA: A Cancer Journal for Clinicians* 70 (1): 7–30.

Song, Zirui, and Katherine Baicker. 2019. "Effect of a Workplace Wellness Program on Employee Health and Economic Outcomes: A Randomized Clinical Trial." *JAMA* 321 (15): 1491–1501.

Starc, Amanda, and Robert J. Town. 2018. "Externalities and Benefit Design in Health Insurance." NBER Working Paper 21783.

Torrance, George W. 1986. "Measurement of Health State Activities for Economic Appraisal." *Journal of Health Economics* 5 (1): 1–30.

US Preventive Services Task Force. 2016. "Final Recommendation Statement: Breast Cancer: Screening." *US Preventive Services Task Force,* January 11. https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/breast-cancer-screening#:~:text=Recommendation%20Summary&text=The%20USPSTF%20recommends%20biennial%20screening,aged%2050%20to%2074%20years.&text=The%20decision%20to%20start%20screening,should%20be%20an%20individual%20one.

US Preventive Services Task Force. 2018a. "Final Recommendation Statement: Prostate Cancer Screening. *US Preventive Services Task Force,* May 8. https://www.uspreventiveservicestaskforce.org/uspstf/recommendation/prostate-cancer-screening.

US Preventive Services Task Force. 2018b. "Grade Definitions." *US Preventive Services Task Force.* https://www.uspreventiveservicestaskforce.org/uspstf/grade-definitions.

Ward, Zachary J., Sara N. Bleich, Angie L. Cradock, Jessica L. Barrett, Catherine M. Giles, Michael W. Long, and Steven L. Gortmaker. 2019. "Projected U.S. State-Level Prevalence of Adult Obesity and Severe Obesity." *New England Journal of Medicine* 381 (25): 2440–50.

Zweifel, Peter, and Friedrich Breyer. 1997. *Health Economics.* Oxford: Oxford University Press.

Mammograms and Mortality: How Has the Evidence Evolved?

Amanda E. Kowalski

A dvocates of preventive care emphasize that it saves money and lives. However, preventive care need not save money (as Newhouse discusses in this issue). Preventive care also need not save lives. In this paper, I analyze evolving evidence on the mortality impact of a common form of preventive care: the mammogram, an X-ray picture of the breast.

The rationale for widespread mammography is that early detection of potentially fatal breast cancers enables earlier and more effective treatment. But there is a potential drawback: mammography can detect some early-stage cancers that will never progress to cause symptoms—a phenomenon often referred to as overdiagnosis. In such cases, the emotional, financial, and physical costs of a cancer diagnosis and any subsequent treatments occur without any corresponding health benefit. Because it is hard to tell which women will be harmed by their cancers, there is a tendency to treat all women as if their cancers will be lethal (Mukherjee 2017). Even if the initial cancer would have never proven life-threatening, exposure to chemotherapy, radiotherapy, and surgery can potentially lead to new conditions, even to new fatal cancers (Praga et al. 2005; Early Breast Cancer Trialists' Collaborative Group 2005).

The possibility of overdiagnosis turns out to be central to guidelines for mammography screening. Prior to 2009, the US Preventive Services Task Force (2002) recommended regular mammography screening for asymptomatic women aged 40 and older. In 2009, the task force revised its guidelines in light of the most

■ *Amanda E. Kowalski is Gail Wilensky Professor of Applied Economics and Public Policy, University of Michigan, Ann Arbor, Michigan. Her email address is aekowals@umich.edu.*

recent follow-up data available from clinical trials on mammography (US Preventive Services Task Force 2009).

The task force's latest guidelines, which are based on a 2016 meta-analysis, reaffirm the 2009 revision. They recommend regular mammography for women aged 50 to 74, but they leave the mammography decision up to individual women in their 40s: "Women who place a higher value on the potential benefit than the potential harms may choose to begin biennial screening between the ages of 40 and 49 years" (Siu 2016). The task force does not provide guidelines for women older than 74 due to insufficient evidence.

In all age groups, mammography has increased dramatically over time in the United States, as shown in Figure 1a. By 2015, 58.3 percent of women aged 40–49, 71.3 percent of women aged 50–65, and 63.3 percent of women aged 65+ reported receiving a mammogram within the past two years. Mammography for women in their 40s has fallen only slightly and gradually since guidelines changed in 2009. Widespread mammography seems embedded in the US health care system, both as a matter of the acculturation of patients and health care providers, and also as a matter of financial incentives: the aggregate annual cost of mammography has been estimated to be $2.1 billion just among US women in their 40s with private health insurance (Kunst et al. 2020).

The US Preventive Services Task Force identifies overdiagnosis as the most important harm that mammograms pose. Though false positives can also pose harm, overdiagnosis is a separate phenomenon. According to the task force definition, overdiagnosis refers to "the diagnosis and treatment of noninvasive and invasive breast cancer that would otherwise not have become a threat to their health, or even apparent, during their lifetime" (Siu 2016).

Overdiagnosis is difficult to identify. If a mammogram catches an early-stage cancer, it is impossible to discern how that cancer would have progressed absent detection, especially for a given individual. However, several types of evidence suggest overdiagnosis of cancer. For example, autopsy studies showing that almost half of older men die with, but not necessarily of, prostate cancer have been important to prostate cancer screening guidelines since the late 1980s (US Preventive Services Task Force 1989). As summarized by Welch and Black (2010), autopsy studies also suggest overdiagnosis of thyroid and breast cancer, and time series evidence suggests overdiagnosis of thyroid, melanoma, and kidney cancer: as diagnoses have risen dramatically with screening, deaths from these types of cancers have been stable.

Deaths from breast cancer have decreased as mammography has increased, but differences in the decreases across geographic areas suggest overdiagnosis. Autier et al. (2010) find that breast cancer mortality for women under age 50 fell from 1989 to 2006 in all 30 European countries that they examine, but they note that the greatest reductions occurred in countries where screening was less common. Using their results, Gøtzsche and Jørgensen (2013) emphasize that the reduction in breast cancer mortality was 49 percent in Denmark but by only 36 percent in Sweden, despite the availability of screening mammography to half of women in

Figure 1

Trends in Mammography and Breast Cancer Incidence in the United States

A: Mammogram within last two years, selected years   B: Breast cancer incidence by stage, 1975–2015
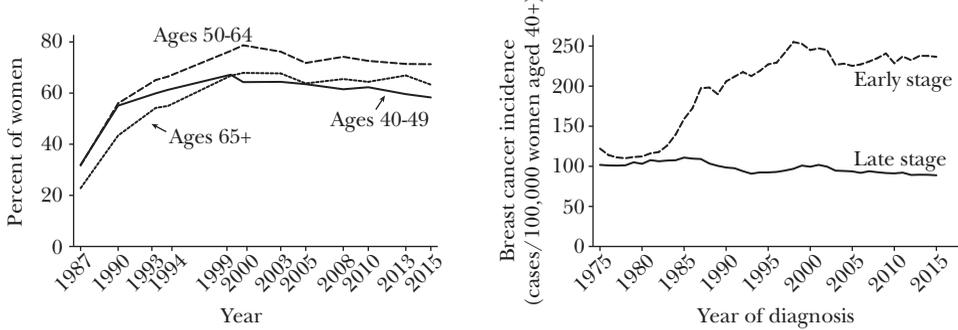
Source: Panel A: National Center for Health Statistics (2011, 2019); Panel B: Author's calculations from Surveillance, Epidemiology, and End Results Program (SEER) (2020).
Note: Incidence is age-adjusted to the 2000 US population. Following Bleyer and Welch (2012), cancer stage is given by the SEER historic stage A variable. Early-stage cancers are in situ or localized. Late-stage cancers are regional or distant.

this age group in Sweden versus none in Denmark. In the United States, breast cancer mortality for women under age 50 fell by an intermediate 39 percent over the same period (author's calculations from Surveillance, Epidemiology, and End Results Program 2019). Difference-in-difference evidence within Denmark for older women aged 55 to 74 shows steeper declines in breast cancer mortality in regions without screening programs from 1997 to 2006 (Jørgensen, Zahl, and Gøtzsche 2010). These findings suggest that decreases in breast cancer mortality are due to factors other than screening, such as decreased use of menopausal hormone therapy (Ravdin et al. 2007) and improved treatments.

Bleyer and Welch (2012) provide an approach that quantifies overdiagnosis from screening. The goal of screening is to prevent early-stage cancers from progressing to a late stage. Therefore, an increase in early-stage cancers without a decrease in late-stage cancers of the same magnitude indicates overdiagnosis, under the assumption that late-stage cancers will be detected regardless of screening. Figure 1b shows that the incidence of early-stage breast cancers in the United States has skyrocketed as the incidence of late-stage breast cancers has declined only slightly. Bleyer and Welch (2012) estimate that in 2008 alone, breast cancer was overdiagnosed for 70,000 women, accounting for 31 percent of all breast cancer diagnoses. Conducting the same exercise using the data in Figure 1b, which replicates their main figure to 2008 and extends it to 2015, I find a similar rate of overdiagnosis in the subsequent seven years.

Evidence of overdiagnosis from randomized controlled trials is the most compelling. Screening detects cancer, so trials should find higher rates of cancer in the intervention arm in the short term. In the long term, if the only impact of screening

is to detect cancer early, rates of cancer should converge in the intervention and control arms as women who experience symptoms have their cancers diagnosed. A failure to converge after a reasonable amount of lead time indicates overdiagnosis. Overdiagnosis became apparent in a lung cancer screening trial based on 16 years of follow-up data (Marcus et al. 2006). Since then, overdiagnosis has become apparent in two studies of mammography trials based on at least 15 years of follow-up data, which imply overdiagnosis rates of 5 to 55 percent depending on the subgroup and base rate (Zackrisson et al. 2006; Baines, To, and Miller 2016).

Overdiagnosis, though undesirable, might be tolerable if screening reduces mortality. The goal of this paper is to analyze how evidence on mortality has evolved over time, across randomized controlled trials of mammography and within one trial. Preventive care such as mammography aims to preempt harm over time, so it is important to consider evolving evidence. I begin with a brief history of the prominent trials, and I present the latest mortality results. These results are the basis for the US Preventive Services Task Force guidelines on mammography screening as well as the basis for guidelines from other public health authorities in the United States and other countries. In a new empirical exercise, I explore the evolution of all-cause mortality relative to breast cancer mortality within a prominent trial, the Canadian National Breast Screening Study, using 20 years of follow-up data. In the conclusion, I offer some responses to the evolving evidence.

A Brief History of Mammography Trials

The 2016 guidelines for mammography screening from the US Preventive Services Task Force are based on a meta-analysis that focuses on mortality impacts obtained from eight large randomized controlled trials (Nelson et al. 2016). All trials combined include over 600,000 women, and individual trials range in size from approximately 39,000 to 160,000. Overall, the meta-analysis finds no statistically significant reduction in all-cause mortality for women in any age group. Some trials even show imprecise increases in all-cause mortality across all age groups or within an age group. Results focused only on breast cancer mortality are slightly more promising. The meta-analysis finds statistically significant but small reductions in breast cancer mortality for women in their 50s and 60s, and it finds imprecise reductions for women aged 39 to 49 and women aged 70 to 74. In this section, I provide a brief overview of the trials and some related evidence on mammography.

The first randomized controlled trial of mammography screening was established in the United States in 1963. Interest in mammograms was growing at the time because breast cancer was the leading cause of cancer death among US women (National Cancer Institute 1979). Before mammograms were used to screen asymptomatic women, they were used to diagnose women with symptoms detected through clinical or self-examination. Large declines in cervical cancer mortality were attributed to pap test screening (Boyes, Knowelden, and Phillips 1973), so there was hope

that mammography screening would reduce breast cancer mortality. Women aged 40-64 who were enrolled in the Health Insurance Plan of Greater New York entered the trial between December 1963 and June 1966. The plan provided medical care on a prepaid basis, with no additional fees for follow-up care. The plan randomly assigned 31,000 women to an intervention arm. Those women were invited for a mammogram and a clinical breast examination, followed by three years of follow-up examinations, which continued even if the women disenrolled from the health insurance plan.

The Health Insurance Plan trial compared the intervention arm to a control arm of 31,000 women who continued to receive care as usual. Preliminary results were published in 1971 based on data through 1969 (Shapiro, Strax, and Venet 1971), which followed almost all women for at least four years and some women for up to six years, depending on their enrollment dates. Of the women invited for screening, about two-thirds appeared for their initial examination, and about two-thirds of those women continued through their third annual examination, resulting in an adherence rate of approximately 45 percent. Of the screened women, biopsies were recommended for 873 women, 545 of which were recommended only on the basis of the clinical exam and not mammography. Of that group, 624 women underwent biopsies, which confirmed breast cancer for 127 of them; for the remaining 497 women, about 80 percent of those who underwent biopsies, the breast cancer diagnosis was a false positive.

Overall, there were 31 breast cancer deaths in the intervention arm and 52 in the control arm. Thus, the "relative risk" of breast cancer mortality in intervention relative to control was about 0.6 ($\approx 31/52$). In absolute terms, there was a less striking reduction of about 7 breast cancer deaths per 10,000 women ($\approx (52\text{-}31)/31,000$)). Almost all this reduction was for women who died at age 50 and older. Among women who died in their 40s, there were 11 breast cancer deaths in intervention and 12 in control. The researchers had followed up intensively on all women with confirmed breast cancer cases to identify breast cancer deaths, but they had only completed their process of gathering information on deaths from all causes among 14 percent of women without confirmed breast cancer cases. In the preliminary data, there were 109 deaths from all causes in intervention but only 99 in control (Shapiro, Strax, and Venet 1971). That is, overall deaths were higher among women invited for screening.

The preliminary results on breast cancer deaths from the Health Insurance Plan trial galvanized support for a massive breast screening program in the United States, the Breast Cancer Detection and Demonstration Project. The project did not involve randomization, although it did facilitate some data collection. Between 1973 and 1980, the project recruited 280,000 women to be screened annually for five years via mammograms and clinical breast exams (Cunningham 1997). The project recruited women as young as 35, and approximately half of participants were under age 50, despite the lack of evidence in favor of mammography for younger women. Even before any long-term data became available, the American Cancer Society started recommending regular mammograms for women aged 50 and older in

1976; it recommended regular mammograms for women aged 40 and older in 1983 (American Cancer Society 2017).

At this point, interest in breast cancer screening programs began to increase around the world, but there was some skepticism given the limited evidence from the Health Insurance Plan Trial. More randomized trials were needed, but given the large numbers of US women who were already receiving mammograms, a follow-up randomized trial would have been practically quite difficult in the United States. A working group convened to review the Breast Cancer Detection Demonstration Project (Summary Report of the Working Group to Review the National Cancer Institute-American Cancer Society Breast Cancer Detection Demonstration Projects 1979) and recommended randomized controlled trials to evaluate questions left unanswered by the Health Insurance Plan trial: for example, should the guidelines recommend mammograms for women in their 40s? And what was the relative contribution of mammography compared to breast examination for women in their 50s?

In 1980, researchers initiated the Canadian National Breast Screening Study to evaluate these and other questions (Miller, Howe, and Wall 1981). Rather than inviting women for screening from a health insurance plan or a population register, this trial conducted a multi-pronged publicity campaign to recruit individual women. Initial media enthusiasm supported recruitment, but over time, high-profile critics fueled public concerns over radiation exposure (Baines 1984). Though overdiagnosis was not central to public concerns, the trial protocol acknowledged concerns related to overdiagnosis as the other main potential hazard from mammography (Miller, Howe, and Wall 1981). The protocol also proposed an approach that has been used to quantify these concerns using data on breast cancer diagnoses from the trial (Miller et al. 2014; Baines, To, and Miller 2016).

The Canadian National Breast Screening Study (CNBSS) enrolled roughly 90,000 women between 1980 and 1985: CNBSS-1 enrolled 50,430 women aged 40 to 49, while CNBSS-2 enrolled 39,405 women aged 50 to 59. These enrollments were in line with power calculations established in the trial protocol (Miller, Howe, and Wall 1981), informed by the latest breast cancer mortality results from the Health Insurance Plan trial (Shapiro 1977). All participants in the CNBSS received a clinical breast exam at enrollment before randomization. Women assigned to the intervention arm received a mammogram during each year of the active study period, which included the enrollment year and the three to four years following enrollment (depending on where and when they enrolled). Women assigned to the control arm received usual care in the community in CNBSS-1 and access to a clinical breast examination during each year of the active study period in CNBSS-2. The trial had the highest adherence rate of all major mammography trials—85 percent (Nelson et al. 2016). For background information on the CNBSS, good starting points include a recent book by Pellerin (2019) and academic publications by the investigators, especially those that report the latest results on mortality and breast cancer incidence 25 years after the first participants enrolled (Miller et al. 2014; Baines, To, and Miller 2016).

Several other international trials began in the 1970s and 1980s. Table 1 summarizes the main trials. It takes information directly from a table within the meta-analysis that informs the current US Preventive Services Task Force guidelines (Nelson et al. 2016). Although Table 1 contains seven rows, Nelson et al. (2016) count CNBSS-1 and CNBSS-2 separately in their discussion of eight trials. Four trials included in Table 1 began in Sweden, in part because Sweden had population registers that could be used to define a trial population by age within a geographic area. Furthermore, it was straightforward to track outcomes because Sweden had a cause of death registry that began in 1951 and a cancer registry that began in 1958. In 1976, the Malmö Mammographic Screening Trial, later known as MMST I, began in Malmö, Sweden. As shown in Table 1, it included approximately 42,000 women. In the following year, 1977, a trial that was more than twice as large, known as the Swedish Two-County Trial, began in Kopparberg and Östergötland. In 1977, the trial in Malmö closed and began enrolling women in an extension trial with the same protocol, known as MMST II. Other trials began in Stockholm in 1981 and Gothenburg in 1982. The Swedish trials have subsequently been pooled for analysis to increase statistical power (Nyström et al. 1993).

Another trial began in Edinburgh, Scotland, in 1978 (Alexander et al. 1999). It recruited approximately 45,000 women from 87 general practices to participate in the first cohort, and it conducted randomization at the practice level. Because of baseline differences between the intervention and control arms, there are substantial concerns about the randomization in the Edinburgh trial. It is therefore not included in the meta-analysis that informs the US Preventive Services Task Force guidelines or in Table 1. Much later, in 1991, another trial began that included participants in Scotland, England, and Wales. That trial is known as the UK Age trial, or simply, the Age trial. Like the Edinburgh trial, the Age trial enrolled women based on their affiliation with general practices, but it conducted randomization at the individual level. An extension of the Age trial, the AgeX trial, began recruiting women aged 47 to 49 and 71 to 73 in 2009. As of late 2018, the AgeX trial has recruited 4 million women, making it over six times larger than the combined size of the trials included in the meta-analysis by Nelson et al. (2016). The trial is powered to detect a 15 percent reduction of breast cancer mortality (Patnick et al. 2018).

Latest Mortality Results across Trials

The meta-analysis by Nelson et al. (2016) that informs the US Preventive Services Task Force mammography guidelines primarily considers the relative risk of breast cancer mortality at the latest available date for each trial. Table 2 reproduces the main results. Within each age group, the table reports results from various studies that provide results for distinct groups within the eight main trials. In all age groups, the relative risk of breast cancer mortality is less than one, indicating that the ratio of the breast cancer mortality rate in intervention to control is less than one. The relative risk is only statistically different from one for women aged 50–59 and 60–69.

Table 1
Summary of Trial Characteristics

Trial Name	Year Trial Began	Setting and Population	Screening Group; Control Group, n*	Method of Randomization	Adherence, %
HIP	1963	New York health plan members aged 40–64	30,239; 30,765	Age- and family size-stratified pairs of women were individually randomized by drawing from a list	46
CNBSS-1 and CNBSS-2	1980	Self-selected partici-pants from 15 centers in Canada aged 40–49 (CNBSS-1), and 50–59 (CNBSS-2)	CNBSS-1: 25,214; 25,216. CNBSS-2: 19,711; 19,694	Individual within blocks stratified by center and 5-year age group after CBE	85
MMST I and MMST II	1976–1978	All women aged 43–69 born between 1908 and 1945 living in Malmö, Sweden	MMST I: 21,088; 21,195. MMST II: 9,581; 8,212	Individual, within birth year	70
Swedish Two-County Trial	1977	Women aged 40–70 from Östergötland and Kopparberg counties in Sweden	77,080; 55,985	Clusters, based on geographic units; blocks designed to be demo-graphically homogenous	84
Stockholm	1981	Residents aged 40–64 from southeast greater Stockholm, Sweden	40,318; 19,943	Individual, by day of month; ratio of screen-ing to control group 2:1	81
Gothenburg	1982	All women aged 39–59 born between 1923 and 1944 living in Gothenburg, Sweden	21,650; 29,961	Cluster, based on day of birth for 1923–1935 cohort (18%), by indi-vidual for 1936–1944 cohort (82%)	75
Age	1991	Women aged 39–41 from 23 National Health Service breast screening units in England, Scotland, and Wales	53,884; 106,956	Individual, stratified by general practitioner group with random number generation 1991–1992; 1992 onward, randomization via Health Authority computer system	57

CBE = clinical breast examination; CNBSS = Canadian National Breast Screening Study; HIP = Health Insurance Plan of New York; MMST = Malmö Mammographic Screening Trial.
* Numbers of participants in screening and control groups vary by publication.

Source: Information taken directly from Nelson et al. (2016) Appendix Table 1.
Note: Some columns from Appendix Table 1 in Nelson et al. (2016) have been omitted, other columns have been included with modified headers and typesetting, and the rows have been reordered to correspond to the narrative in this paper. Although this table contains seven rows, Nelson et al. (2016) count CNBSS-1 and CNBSS-2 separately in their discussion of eight trials.

Table 2

Effects on Breast Cancer Mortality at Latest Available Date

Author (Year)	Trial Name	Mean Follow-up, Years	Relative Risk (95% CI)
Women aged 39–49 years			
Habbema et al. (1986)	HIP	14.0	0.75 (0.53–1.05)
Miller et al. (2014)	CNBSS-1	21.9	1.04 (0.87–1.24)
Nyström et al. (2002)*	MMST I	18.2	0.74 (0.42–1.29)
Nyström et al. (2002)*	MMST II	11.2	0.64 (0.39–1.06)
Tabar et al. (1995)	Östergötland	12.5	1.02 (0.52–1.99)
Tabar et al. (1995)	Kopparberg	12.5	0.73 (0.37–1.41)
Nyström et al. (2002)*	Stockholm	14.3	1.52 (0.80–2.88)
Bjurstam et al. (2003)	Gothenburg	13.8	0.69 (0.45–1.05)
Moss et al. (2015)	Age	17.5	0.93 (0.80–1.09)
Overall ($I^2 = 25\%$; $p = 0.230$)			0.92 (0.75–1.02)
Women aged 50–59 years			
Habbema et al. (1986)	HIP	14.0	0.83 (0.61–1.13)
Miller et al. (2014)	CNBSS-2	21.9	0.94 (0.78–1.13)
Nyström et al. (2002)*	MMST I	18.1	0.98 (0.75–1.29)
Tabár et al. (1995)	Östergötland	12.5	0.85 (0.52–1.38)
Tabár et al. (1995)	Kopparberg	12.5	0.48 (0.29–0.77)
Nyström et al. (2002)*	Stockholm	13.7	0.56 (0.32–0.97)
Bjurstam et al. (2003)	Gothenburg	13.8	0.83 (0.60–1.15)
Overall ($I^2 = 38.0\%$; $p = 0.139$)			0.86 (0.68–0.97)
Women aged 60–69 years			
Habbema et al. (1986)	HIP	14.0	0.85 (0.48–1.47)
Nyström et al. (2002)*	MMST I	15.5	0.64 (0.45–0.92)
Tabár et al. (1995)	Östergötland	12.5	0.62 (0.43–0.91)
Tabár et al. (1995)	Kopparberg	12.5	0.58 (0.35–0.96)
Nyström et al. (2002)*	Stockholm	13.1	0.94 (0.46–2.02)
Overall ($I^2 = 0.0\%$; $p = 0.739$)			0.67 (0.54–0.83)
Women aged 70–74 years			
Nyström et al. (2002)*	MMST I	13.6	0.98 (0.15–6.60)
Tabár et al. (1995)	Östergötland	12.5	0.82 (0.43–1.58)
Tabár et al. (1995)	Kopparberg	12.5	0.76 (0.42–1.36)
Overall ($I^2 = 0.0\%$; $p = 0.962$)			0.80 (0.51–1.28)

Meta-analysis of trials using the longest follow-up times available. CNBSS = Canadian National Breast Screening Study; HIP = Health Insurance Plan of New York; MMST = Malmö Mammographic Screening Trial.

*Used short case accrual.

Source: Information taken directly from Nelson et al. (2016) Figure 1.

Note: CI = confidence interval. The columns and panels from Figure 1 in Nelson et al. (2016) have been included with modified headers and typesetting, and the rows have been reordered to correspond to the narrative in this paper. The full version of Figure 1 in Nelson et al. (2016) also contains plots of relative risk and the respective 95 percent CI. I^2 is a measure of inconsistency across studies, which ranges from 0–100 percent.

Nelson et al. (2016) note that the implied reduction in breast cancer mortality is small in absolute terms. Screening 10,000 women aged 50–59 over 10 years prevents 7.7 breast cancer deaths, and screening the same number of women aged 60–69 over 10 years prevents 21.3 (Nelson et al. 2015). The number of breast cancer

deaths avoided due to mammography screening seems considerably lower than the number of cause-specific deaths avoided due to screening for other cancers. As a point of comparison, the US Preventive Services Task Force guidelines on colorectal cancer screening report that it prevents 200 to 240 colorectal cancer deaths per 10,000 people aged 50 to 70 (Bibbins-Domingo et al. 2016).

Deaths attributed to a particular type of cancer may not capture the full mortality impact of screening. Cause of death coding can be a subjective enterprise. For example, mammography screening can increase the probability of a breast cancer diagnosis, which can increase the probability that a death is coded as a breast cancer death. Through this mechanism, mammography trial estimates can be biased against showing reductions in breast cancer mortality. However, mammography can also lead to overdiagnosis and other collateral harms, which may or may not be captured in breast cancer mortality. The overall direction of the bias is unclear. Thus, it seems useful to examine all-cause mortality to capture a wider range of potential benefits and harms from mammograms.

The meta-analysis that informs the US Preventive Services Task Force mammography guidelines includes an analysis of all-cause mortality in a technical document (Nelson et al. 2015). Table 3 reproduces the results here. Combined analysis of all trials indicates a relative risk of 0.99 for women aged 39 to 49, 0.97 for women aged 60 to 69, and 0.98 for women aged 70 to 74. None of these relative risks are significantly different from one at the 5 percent level. Furthermore, the relative risk is 1.02 for women aged 50 to 59, indicating net harm. Several individual studies also report relative risks greater than one. All of these relative risks are imprecise and should thus be interpreted with caution. Imprecision aside, relative risks greater than one are striking. Even if a trial shows overdiagnosis in terms of breast cancer incidence, overdiagnosis could still be innocuous in terms of all-cause mortality.

It should also be noted that, for many trials, the mean follow-up for all-cause mortality reported in Table 3 is much shorter than the mean follow-up for breast cancer mortality reported in Table 2. The lack of later follow-up is concerning, given that reductions in mortality due to prevention of late-stage cancers might be expected to happen relatively sooner in the follow-up period, while negative effects of overdiagnosis might manifest relatively later. In one trial, the Canadian National Breast Screening Study, results in terms of all-cause mortality have tended to weaken the case for mammography over time (Miller et al. 1992a,b, 2014). Later in this paper, I will offer some new evidence on the time profile of impacts on breast cancer mortality and all-cause mortality using the data from this trial.

It is hard to understate the controversy surrounding the results from the mammography trials. The idea that finding small, treatable cancers will save lives by stopping them from growing into larger malignant cancers is appealingly simple. Furthermore, the mammography trials began in an era in which sexism and paternalism toward women were much more overt than they are today. Spurred by the Breast Cancer Detection Demonstration Project, manufacturers of mammography equipment were eager to expand their market. Billboard signs implored, "If you don't get a mammogram, you need more than your breasts examined" (Pellerin 2019).

Table 3
Effects on All-Cause Mortality at Latest Available Date

Author (Year)	Trial Name	Mean Follow-up, Years	Relative Risk (95% CI)
Women aged 39–49 years			
Miller et al. (2002)	CNBSS-1	13.0	1.00 (0.87–1.15)
Nyström et al. (2002)	Malmö II	9.1	1.03 (0.89–1.20)
Tabár et al. (1989)	Östergötland	7.9	0.93 (0.76–1.12)
Tabár et al. (1989)	Kopparberg	7.9	1.33 (1.01–1.77)
Frisell et al. (1997)	Stockholm	11.0	1.12 (0.55–2.41)
Bjurstam et al. (1997)	Gothenburg	10.0	0.98 (0.86–1.12)
Moss et al. (2006)	Age	10.7	0.97 (0.89–1.04)
Overall ($I^2 = 0.0\%$; $p = 0.478$)			0.99 (0.94–1.06)
Women aged 50–59 years			
Miller et al. (2000)	CNBSS-2	13.0	1.06 (0.96–1.18)
Tabár et al. (1989)	Östergötland	7.9	0.98 (0.87–1.11)
Tabár et al. (1989)	Kopparberg	7.9	1.00 (0.86–1.17)
Overall ($I^2 = 0.0\%$; $p = 0.588$)			1.02 (0.94–1.10)
Women aged 60–69 years			
Tabár et al. (1989)	Östergötland	7.9	0.98 (0.91–1.05)
Tabár et al. (1989)	Kopparberg	7.9	0.95 (0.87–1.04)
Overall ($I^2 = 0.0\%$; $p = 0.650$)			0.97 (0.90–1.04)
Women aged 70–74 years			
Tabár et al. (1989)	Östergötland	7.9	0.93 (0.87–1.01)
Tabár et al. (1989)	Kopparberg	7.9	1.05 (0.95–1.15)
Overall ($I^2 = 72.4\%$; $p = 0.057$)			0.98 (0.86–1.14)
Combined ages			
Aron and Prorok (1986)	HIP	10	0.99 (0.93–1.05)
Miller et al. (2014)	CNBSS-1 & 2	25	1.02 (0.98–1.06)
Nyström et al. (2002)	Malmö I	19.2	0.99 (0.97–1.01)
Nyström et al. (2002)	Malmö II	9.1	1.03 (0.89–1.20)
Nyström et al. (2002)	Östergötland	17.2	0.98 (0.95–1.01)
Nyström et al. (2002)	Stockholm	14.7	0.99 (0.95–1.03)
Nyström et al. (2002)	Gothenburg	13.2	0.94 (0.88–1.00)
Moss et al. (2006)	Age	10.7	0.97 (0.89–1.04)
Overall ($I^2 = 0.0\%$; $p = 0.577$)			0.99 (0.97–1.003)

CI = confidence interval; CNBSS = Canadian National Breast Screening Study; HIP = Health Insurance Plan of Greater New York.

Source: Information taken directly from Nelson et al. (2015) Figures 7 and 8.
Note: Malmö = Malmö Mammographic Screening Trial. The top four panels take information from Nelson et al. (2015) Figure 8, and the bottom panel takes information from Figure 7. The columns and panels from these figures have been included with modified headers and typesetting, and the rows have been reordered to correspond to the narrative in this paper. The full versions of these figures also contain plots of relative risk and the respective 95 percent CI. The ages included for each trial in the "Combined ages" panel can be found in the "Age, *year*" column of Nelson et al. (2015) Figure 7, which has been omitted here. I^2 is a measure of inconsistency across studies, which ranges from 0–100 percent.

For advocates of mammography, it was thus disappointing when the evidence from the randomized controlled trials was underwhelming. Some began to update their thinking on the value of mammography. Others challenged the methodologies of the trials themselves, particularly whether the randomization had been adequate. The meta-analysis that informs the US Preventive Services Task Force guidelines rated the randomization of all of the trials in Table 1 as "fair" quality but raised various concerns about specific studies in terms like: "Generally effective randomization and comparable groups are assembled initially, but some question remains whether some, although not major, differences occurred in follow-up," "Important differential loss to follow-up or overall high loss to follow-up; adherence <80%," "Numbers of participants unclear," and "Did not maintain comparable groups (includes attrition, crossovers, adherence, contamination)." For comparison, a Cochrane review deemed most of the trials, including the second part of the Malmö trial (MMST II) to be "suboptimally" randomized, but it also deemed the Canadian National Breast Screening Study, the first part of the Malmö trial (MMST I), and the Age trial to be "adequately randomized" (Gøtzsche and Jørgensen 2013).

Trial Results Inform Mammography Guidelines

The evidence from mammography trials has informed mammography guidelines in many countries. For example, Swedish trials are the only trials in Table 2 that show statistically significant decreases in breast cancer mortality in some age groups. Swedish national guidelines recommend mammograms for women aged 40 to 74 (Ebell, Thai, and Royalty 2018).

In contrast, Canadian national guidelines "recommend not screening" with mammography for women aged 40 to 49 (Klarenbach et al. 2018) but "recommend screening with mammography" for women aged 50 to 74. As summarized in Table 2, the most recent mortality results from the Canadian National Breast Screening Study (Miller et al. 2014) show an imprecise increase in breast cancer mortality for women in their 40s and an imprecise decrease in breast cancer mortality for women in their 50s. The most recent breast cancer incidence results indicate overdiagnosis (Baines, To, and Miller 2016).

Many other high income countries, including Australia, France, Switzerland, and the United Kingdom, do not recommend mammography for women in their 40s, and they also do not recommend against it as Canadian guidelines do (Ebell, Thai, and Royalty 2018). However, the Swiss Medical Board recommended steps to limit screening programs in 2014 (Biller-Andorno and Jüni 2014). In 2016, the French Minister of Health released results of an independent review that recommended that the national screening program end or undergo radical reforms (Barratt, Jørgensen, and Autier 2018).

Within the United States, guidelines vary but tend to make selective recommendations similar to those of the US Preventive Services Task Force (CDC 2020). For women in their 40s, the US Preventive Services Task Force, the American Academy

of Family Physicians, and the American College of Physicians all leave the mammography decision up to individual women and their doctors, as does the American Cancer Society for women aged 40 to 44. In contrast, the American College of Radiology recommends regular mammograms for women in their 40s.

Evolution of Mortality Results within a Trial

The medical literature has commonly evaluated randomized controlled trials of mammography by using breast cancer mortality as the primary outcome, and it commonly examines the outcome only at the latest available date. The meta-analysis that informs the US Preventive Services Task Force mammography guidelines follows these practices too. However, breast cancer mortality may or may not capture mortality impacts of mammography that occur through overdiagnosis (and perhaps false positives). Indeed, some in the medical literature argue for a focus on all-cause mortality (Black, Haggstrom, and Welch 2002). In addition, measuring the outcome of a mammography trial at the latest available date does not reveal how benefits and harms of mammography may develop over different timeframes.

In this section, I reanalyze data from the Canadian National Breast Screening Study (CNBSS 2015), the trial with the longest follow-up length of all trials considered by the Nelson et al. (2016) meta-analysis, with an emphasis on how breast cancer mortality and all-cause mortality evolve over time. Each intermediate result that I examine reflects a fixed follow-up length since enrollment. In contrast, the CNBSS investigators examine results at the latest available calendar date, which is not quite the same thing. Because participants enrolled in the CNBSS over a period of five years, the most recent CNBSS results reflect follow-up lengths that vary from 20 to 25 years across participants. The practice of aggregating information across multiple follow-up lengths is common and reasonable because it preserves balance between intervention and control and incorporates all available information. For example, the meta-analysis that informs the US Preventive Services Task Force guidelines reports an estimate from the latest available date for each trial (Nelson et al. 2016) and reports the mean follow-up length, per Table 2. However, this practice makes trends difficult to interpret.

Mammography trials often reassess results as data accumulates and the latest available date advances (Miller et al. 1992a,b, 2000, 2002, 2014), and some research in the medical literature has emphasized systematic analysis of cause-specific mortality over time (Miettinen et al. 2002; Hanley 2005, 2010, 2011; Hanley et al. 2013; Liu et al. 2015). Here, I will show that a systematic assessment of all-cause mortality over time reveals additional information.

As a starting point, I examine breast cancer mortality and all-cause mortality 20 years after enrollment in this study, which is the maximum follow-up length available for all participants. Following the latest mortality results published by the Canadian National Breast Screening Study investigators (Miller et al. 2014), I begin by pooling results for women of all ages. The first row of Table 4 shows that

Table 4

Excess Breast Cancer Mortality and All-Cause Mortality Rates in Intervention 20 Years after Enrollment in the Canadian National Breast Screening Study

	Intervention (1)	Control (2)	Intervention – Control (1) – (2)
Breast dancer deaths (per 100,000)	904	924	−20
			(65)
All-cause deaths (per 100,000)	7,969	7,880	89
			(175)
Observations	44,925	44,910	

Note: Years after enrollment are constructed for each subject based on the exact calendar date of enrollment. Standard errors in parentheses are calculated as the standard deviation of the point estimates obtained in 200 bootstrap samples. Subjects aged 40–59 at enrollment are included.

the intervention arm experienced 20 fewer deaths from breast cancer per 100,000 women than the control arm. This imprecise reduction in breast cancer mortality in the intervention arm is consistent with the negative and statistically insignificant point estimate published by the CNBSS investigators at the latest available date (Miller et al. 2014), and it is replicated in the last column of the second panel of Table A.1 in the online Appendix available with the paper at the *JEP* website.

I next examine the excess all-cause mortality rate in the intervention arm 20 years after enrollment. The second row of Table 4 shows that the intervention arm experienced 89 more all-cause deaths per 100,000 women than the control arm. Again, this estimate is not statistically different from zero, and it is consistent with the positive but statistically insignificant point estimate published by the Canadian National Breast Screening Study investigators at the latest available date (Miller et al. 2014), replicated in the last column of the second panel of Table A.1 in the online Appendix.

The next step is to look at the evolution of breast cancer mortality over time. I start by constructing point estimates analogous to the point estimate reported in the first row of Table 4 for each intermediate annual follow-up length, starting with the enrollment year as year 0. I plot the results in Figure 2. All but two of the point estimates are statistically indistinguishable from zero. To investigate the existence of a trend, I plot the line of best fit across all point estimates. I obtain the statistical significance of the trend using the equivalent panel regression to take into account that the point estimates as well as the trend are estimated. The trend line appears flat with a slope that is not statistically different from zero. Based on these results, it is not surprising that the literature, which focuses on breast cancer mortality as the primary outcome at various points in time, does not identify a compelling trend in breast cancer mortality as follow-up length increases.
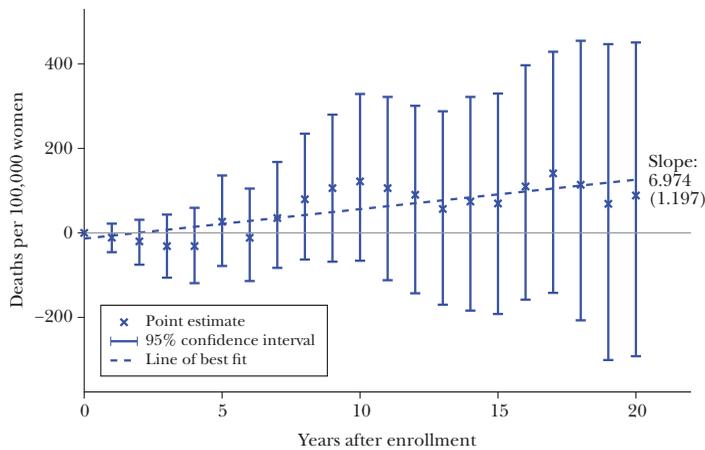
I next carry out the same approach with all-cause mortality as the outcome of interest. As shown in Figure 3, through 20 years after enrollment, point estimates are not statistically different from zero at any follow-up length. However, a pattern

Figure 2

Trend in Excess Breast Cancer Mortality Rate in Intervention in the Canadian National Breast Screening Study

Note: Years after enrollment are constructed for each subject based on the exact calendar date of enrollment. 95 percent confidence intervals in each year after enrollment are obtained from the same 200 bootstrap samples. The standard error on the slope of the line of best fit, in parentheses, is calculated from a panel regression, block bootstrapped by year after enrollment, which takes into account that the point estimates and the line of best fit are estimated. Subjects aged 40–59 at enrollment are included.

Figure 3

Trend in Excess All-Cause Mortality Rate in Intervention in the Canadian National Breast Screening Study

Note: Years after enrollment are constructed for each subject based on the exact calendar date of enrollment. 95 percent confidence intervals in each year after enrollment are obtained from the same 200 bootstrap samples. The standard error on the slope of the line of best fit, in parentheses, is calculated from a panel regression, block bootstrapped by year after enrollment, which takes into account that the point estimates and the line of best fit are estimated. Subjects aged 40–59 at enrollment are included.

emerges over time: point estimates in the first four years after enrollment are negative. Year five is the first positive point estimate, and the point estimates are always positive after year six. Although these point estimates are individually imprecise, they are consistent with a hypothesis that life-saving benefits of mammograms outweigh their collateral harms at first, but as more time passes, collateral harms rise.

I plot the line of best fit across all point estimates, which I obtain from an equivalent panel regression, in Figure 3. A pronounced upward-sloping and statistically significant trend is visible in the all-cause mortality results. With each additional year that passes after enrollment, an additional seven excess deaths per 100,000 women become apparent among intervention-arm participants relative to control-arm participants. To put this trend in perspective, annual road traffic deaths in the United States are 12.4 per 100,000 (World Health Organization 2018).

In Figures A.1 and A.2 in the online Appendix, I show that the positive trend is larger in magnitude for women aged 40–49 at enrollment (a slope of around 8 deaths per 100,000 women per year) than for women aged 50–59 at enrollment (a slope of around 6 deaths per 100,000 women per year). This difference by age group is unsurprising given the weaker evidence on the effect on breast cancer mortality for younger women and the corresponding difference in the US Preventive Services Task Force guidelines for women in their 40s.

There are many mechanisms that could explain the trend in all-cause mortality. For example, breast cancer treatments could have mortality impacts that are not captured by breast cancer mortality. Randomized controlled trials of radiotherapy show that it increases lung and esophageal cancer mortality among women treated for early-stage breast cancer (Early Breast Cancer Trialists' Collaborative Group 2005). I add lung and esophageal cancer deaths to breast cancer deaths and estimate the trend. From these three causes combined, I find that an additional 3.6 excess deaths per 100,000 women become apparent in each additional year after enrollment. This trend is statistically significant, and it can explain more than half of the trend in all-cause mortality (results shown in Figure A.3 in the online Appendix). Perhaps deaths from other types of cancer can explain some of the rest of the difference. However, other cause-specific measures of mortality are subject to similar limitations as breast cancer mortality.

The trend in all-cause mortality is especially striking given that mammography likely increased in the control arm over time. Although the investigators did not collect complete data on mammograms received after the active study period, Canadian breast screening programs began in the late 1980s and 1990s, so it seems likely that some women in the control arm began receiving regular mammography at some point within 20 years of enrollment (Baines, To, Miller 2016). Convergence in mammography behavior should attenuate mortality differences between the intervention and control arms over time—which makes the positive trend in excess all-cause mortality in the intervention arm all the more surprising. Convergence in mammography behavior also implies that the results speak to whether mammography should begin at younger ages, say, in a woman's 40s rather than in her 50s, in line with the most recent change in the US Preventive Services Task Force guidelines.

These findings should be interpreted with caution. The Canadian National Breast Screening Study began decades ago. In any of the mammography trials, factors such as mammogram technology and breast cancer treatment could have changed over time, so past trends could differ from future trends in ways that are difficult to assess. For instance, technological improvements that allow mammograms to identify smaller tumors could worsen the mortality impact of overdiagnosis if those smaller tumors are less likely to advance to late-stage cancers. In contrast, innovations that make breast cancer treatment milder could alleviate the mortality impact of overdiagnosis.

Despite these challenges, evolving results from mammography trials represent the best available evidence. Examination of the trend in all-cause mortality reveals that the tradeoff between the harms and benefits of mammography has been shifting toward harms over time. It also offers a prediction of how the evidence will continue to evolve, which is useful since organizations such as the US Preventive Services Task Force update guidelines over time.

Responses to Evolving Evidence

One response to the evolving evidence on mortality from mammography trials is to incorporate it into new policies. Section 2713 of the Affordable Care Act (2010) ties health insurance coverage for preventive services to current and future guidelines of the US Preventive Services Task Force. However, it makes an explicit exception for mammography, allowing coverage for annual mammograms for women in their 40s, despite current guidelines.

Another response is to reverse existing policies that are at odds with current guidelines, such as policies that lead to over-utilization of mammography among women older than 74. For example, California state auditors implemented a policy in the late 1980s that required many elderly women enrolled in a particular health plan to receive at least one mammogram. The mean age of the women screened between 1995 and 1997 as a result of the policy was 81. Unsurprisingly, 17 percent of these women experienced significant burdens from mammography (Walter, Eng, and Covinsky 2001). This policy was later reversed, but the issue of screening older women persists. One doctor told Kaiser Health News that doctors continue to screen older women because of fears of lawsuits, health system bonuses for high screening rates, and because "doing less can be perceived as a lack of caring or as ageism . . . It can be uncomfortable for a physician to explain why doing less is more" (Szabo 2017).

Even if it is difficult to limit mammography, medical practice can respond to evolving evidence on mammography by reducing or at least postponing the use of potentially unnecessary treatments, especially treatments for early-stage cancers identified with mammography. The *New York Times* described the results of a recent clinical trial of breast cancer treatment as "good news for women with breast cancer: many don't need chemo" (as reported in Grady 2018). The trial found that the

addition of chemotherapy to endocrine therapy does not increase invasive disease-free survival for women with certain breast cancer diagnoses, but it has some benefit for women aged 50 or younger whose breast cancers are initially deemed more likely to recur (Sparano et al. 2018).

Yet another response to the evolving evidence on mammography is to produce more targeted evidence. In that vein, it is instructive to learn from responses to experimental evidence in development economics. As Esther Duflo (2020) discusses in her Nobel lecture, initial responses from the microfinance industry were critical of experimental evidence that contradicted the perception of microfinance as a panacea against poverty. However, experimental evidence was never meant to pass a categorical judgment on microfinance; rather, it aimed to uncover contexts in which microfinance may work well. As with microfinance, the underwhelming evidence on benefits of mammography should not motivate a wholesale rejection of the practice, but rather it should motivate research aimed at uncovering the contexts in which mammograms may provide benefits.

Existing and new methods can be used to predict the characteristics of women for whom mammography is most beneficial within trial data. In my ongoing research, I propose an approach to quantify the number of individuals in a trial who would be harmed by an intervention, such as access to mammography, even if the trial does not show harm on average (Kowalski 2020b). Advances in machine learning can then be applied to determine which women are most likely to benefit. With this knowledge in hand, examination of which women select into mammography (as in Kim and Lee 2017; Einav et al. 2020), and how these selection patterns translate into heterogeneous treatment effects of mammography (as in Kowalski 2020a) can help policymakers craft better targeted policies. There has already been careful work that examines the impact of policies on mammography (Mehta et al. 2015; Bitler and Carpenter 2016; Kadiyala and Strumpf 2016; Lu and Slusky 2016; Buchmueller and Goldzahl 2018; Myerson et al. 2020). Such work can be extended to examine the targeting of mammography in light of evidence on mortality from clinical trials.

References

Affordable Care Act. 2010. *The Patient Protection and Affordable Care Act, Sec. 2713, Coverage of Preventive Services.* Washington, DC: US Government Printing Office.

Alexander, F., T. Anderson, H. Brown, A. Forrest, W. Hepburn, A. Kirkpatrick, B. Muir, R. Prescott, and A. Smith. 1999. "14 Years of Follow-up from the Edinburgh Randomised Trial of Breast-Cancer Screening." *The Lancet* 353 (9168): 1903–08.

American Cancer Society. 2017. History of ACS Recommendations for the Early Detection of Cancer in People without Symptoms. https://www. cancer.org/healthy/find-cancer-early/cancer-screening-guidelines/ chronological-history-of-acs-recommendations.html. (accessed March 12, 2018).

Aron, Joan L., and Philip C. Prorok. 1986. "An Analysis of the Mortality Effect in a Breast Cancer Screening Study." *International Journal Epidemiology* 15(1): 36–43.

Autier, Philippe, Mathieu Boniol, Carlo LaVecchia, Lars Vatten, Anna Gavin, Clarisse Héry, and Mary Heanue. 2010. "Disparities in Breast Cancer Mortality Trends between 30 European Countries: Retrospective Trend Analysis of WHO Mortality Database." *Bmj* 341 (3620).

Baines, C. J. 1984. "Impediments to Recruitment in the Canadian National Breast Screening Study: Response and Resolution." *Controlled Clinical Trials* 5 (2): 129–40.

Baines, Cornelia J., Teresa To, and Anothony B. Miller. 2016. "Revised Estimates of Overdiagnosis from the Canadian National Breast Screening Study." *Preventive Medicine* 90: 66–71.

Barratt, Alexandra, Karsten Juhl Jørgensen, and Philippe Autier. 2018. "Reform of the National Screening Mammography Program in France." *JAMA Internal Medicine* 178 (2): 177–78.

Bibbins-Domingo, Kirsten, David C. Grossman, Susan J. Curry, Karina W. Davidson, John W. Epling Jr., Francisco A.R. García, Matthew W. Gillman et al. 2016. "Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement." *JAMA* 315 (23): 2564–75.

Biller-Andorno, Nikola, and Peter Jüni. 2014. "Abolishing Mammography Screening Programs? A View from the Swiss Medical Board." *New England Journal of Medicine* 370: 1965–67.

Bitler, Marianne P. and Christopher S. Carpenter. 2016. "Health Insurance Mandates, Mammography, and Breast Cancer Diagnoses." *American Economic Journal: Economic Policy* 8 (3): 39–68.

Bjurstam, Nils, Lena Björneld, Stephen W. Duffy, Teresa C. Smith, Erling Cahlin, Olof Eriksson, Lars-Olof Hafström et al. 1997. "The Gothenburg Breast Screening Trial: First Results on Mortality, Incidence, and Mode of Detection for Women Ages 39–49 Years at Randomization." *Cancer* 80(11): 2091–99.

Bjurstam, Nils, Lena Björneld, Jane Warwick, Evis Sala, Stephen W. Duffy, Lennarth Nyström, Neil Walker et al. 2003. "The Gothenburg Breast Screening Trial." *Cancer* 97 (10): 2387–96

Black, William C., David A. Haggstrom, and H. Gilbert Welch. 2002. "All-Cause Mortality in Randomized Trials of Cancer Screening." *Journal of the National Cancer Institute* 94 (3): 167–73.

Bleyer, Archie, and H. Gilbert Welch. 2012. "Effect of Three Decades of Screening Mammography on Breast-Cancer Incidence." *The New England Journal of Medicine* 367: 1998–2005.

Boyes, D., J. Knowelden, and A. Phillips. 1973. "The Evaluation of Cancer Control Measures." *British Journal of Cancer* 28: 105–07.

Buchmueller, Thomas C. and Léontine Goldzahl. 2018. "The Effect of Organized Breast Cancer Screening on Mammography Use: Evidence from France." *Health Economics* 27 (12): 1963–80.

CDC. 2020. Breast Cancer Screening Guidelines for Women. https://www.cdc.gov/ cancer/ breast/pdf/ breast-cancer-screening-guidelines-508.pdf (accessed September 24, 2020).

CNBSS. 2015. "Canadian National Breast Screening Study: Cumulative Data 1980-2005." University of Toronto.

Cunningham, Myles P. 1997. "The Breast Cancer Detection Demonstration Project 25 Years Later." *CA: A Cancer Journal for Clinicians* 47 (3): 131–33.

Duflo, Esther. 2020. "Field Experiments and the Practice of Policy." *American Economic Review* 110 (7): 1952–73.

Early Breast Cancer Trialists' Collaborative Group. 2005. "Effects of Radiotherapy and of Differences in the Extent of Surgery for Early Breast Cancer on Local Recurrence and 15-Year Survival: An Overview of the Randomised Trials." *The Lancet* 366 (9503): 2087–2106.

Ebell, Mark H., Thuy Nhu Thai, and Kyle J. Royalty. 2018. "Cancer Screening Recommendations: An International Comparison of High Income Countries." *Public Health Reviews* 39 (1).

Einav, Liran, Amy Finkelstein, Tamar Oostrom, Abigail J. Ostriker, and Heidi Williams. 2020. "Screening

and Selection: The Case of Mammograms." *American Economic Review* 110 (12): 3836–70.

Frisell, Jan, Elisabet Lidbrink, Lars Hellström, and Lars-Erik Rutqvist. 1997. "Followup after 11 years–Update of Mortality Results in the Stockholm Mammographic Screening Trial." *Breast Cancer Research and Treatment* 45 (3): 263–70.

Gøtzsche, Peter C., and Karsten Juhl Jørgensen. 2013. "Screening for Breast Cancer with Mammography." *Cochrane Database of Systematic Reviews* 6: 1–73.

Grady, Denise. 2018. "Good News for Women with Breast Cancer: Many Don't Need Chemo." *New York Times*, June 3. https://www.nytimes.com/2018/06/03/health/breast-cancer-chemo.html.

Habbema, J. D. F., Gerril J. van Oortmarssen, Dick J. van Putten, Jacobus T. Lubbe, and Paul J. van der Maas. 1986. "Age-Specific Reduction in Breast Cancer Mortality by Screening: An Analysis of the Results of the Health Insurance Plan of Greater New York Study." *Journal of the National Cancer Institute* 77 (2): 317–20.

Hanley, James A. 2005. "Analysis of Mortality Data from Cancer Screening Studies: Looking in the Right Window." *Epidemiology* 16 (6): 786–90.

Hanley, James A. 2010. "Mortality Reductions Produced by Sustained Prostate Cancer Screening Have Been Underestimated." *Journal of Medical Screening* 17 (3): 147–51.

Hanley, James A. 2011. "Measuring Mortality Reductions in Cancer Screening Trials." *Epidemiologic Reviews* 33 (1): 36–45.

Hanley, James A., Maurice McGregor, Zhihui Liu, Erin C. Strumpf, and Nandini Dendukuri. 2013. "Measuring the Mortality Impact of Breast Cancer Screening." *Canadian Journal of Public Health* 104: 437–42.

Jørgensen, Karsten J., Per-Henrik Zahl, and Peter C. Gøtzsche. 2010. "Breast Cancer Mortality in Organised Mammography Screening in Denmark: Comparative Study." *BMJ* 340 (1241).

Kadiyala, Srikanth, and Erin Strumpf. 2016. "How Effective Is Population-Based Cancer Screening? Regression Discontinuity Estimates from the US Guideline Screening Initiation Ages." *Forum for Health Economics & Policy* 19 (1): 87–139.

Kim, Hyuncheol Bryant, and Sun-mi Lee. 2017. "When Public Health Intervention Is Not Successful: Cost Sharing, Crowd-Out, and Selection in Korea's National Cancer Screening Program." *Journal of Health Economics* 53: 100–116.

Klarenbach, Scott, Nicki Sims-Jones, Gabriela Lewin, Harminder Singh, Guylène Thériault, Marcello Tonelli, Marion Doull, Susan Courage, Alejandra Jaramillo Garcia, and Brett D. Thombs. 2018. "Recommendations on Screening for Breast Cancer in Women Aged 40–74 Years Who Are Not at Increased Risk for Breast Cancer." *Canadian Medical Association Journal* 190 (49): 1441–51.

Kowalski, Amanda E. 2020a. "Behavior within a Clinical Trial and Implications for Mammography Guidelines. NBER Working Paper 25049.

Kowalski, Amanda E. 2020b. "Counting Defiers: Examples from Health Care." https://arxiv.org/abs/1912.06739.

Kunst, Natalia, Jessica B. Long, Xiao Xu, Susan H. Busch, Kelly A. Kyanko, Ilana B. Richman, and Cary P. Gross. 2020. "Use and Costs of Breast Cancer Screening for Women in Their 40s in a US Population with Private Insurance." *JAMA Internal Medicine* 180 (5): 799–801.

Liu, Zhihui, James A. Hanley, Olli Saarela, and Nandini Dendukuri. 2015. "A Conditional Approach to Measure Mortality Reductions Due to Cancer Screening." *International Statistical Review* 83 (3): 493–510.

Lu, Yao, and David J. Slusky. 2016. "The Impact of Women's Health Clinic Closures on Preventive Care." *American Economic Journal: Applied Economics* 8 (3): 100–124.

Marcus, Pamela M., Erik J. Bergstralh, Mark H. Zweig, Ann Harris, Kenneth P. Offord, and Robert S. Fontana. 2006. "Extended Lung Cancer Incidence Follow-up in the Mayo Lung Project and Overdiagnosis." *Journal of the National Cancer Institute* 98 (11): 748–56.

Mehta, Shivan J., Daniel Polsky, Jingsan Zhu, James D. Lewis, Jonathan T. Kolstad, George Loewenstein, and Kevin G. Volpp. 2015. "ACA-Mandated Elimination of Cost Sharing for Preventive Screening Has Had Limited Early Impact." *The American Journal of Managed Care* 21 (7): 511–17.

Miettinen, Olli S., Claudia I. Henschke, Mark W. Pasmantier, James P. Smith, Daniel M. Libby, and David F. Yankelevitz. 2002. "Mammographic Screening: No Reliable Supporting Evidence? *The Lancet* 359 (9304): 404–05.

Miller, A., G. Howe, and C. Wall. 1981. "The National Study of Breast Cancer Screening." *Clinical & Investigative Medicine* 4 (3/4): 227–58.

Miller, A. B., C. J. Baines, T. To, and C. Wall. 1992a. "Canadian National Breast Screening Study: 1.

Breast Cancer Detection and Death Rates among Women Aged 40 to 49 years." *Canadian Medical Association Journal* 147 (10): 1459–76.

Miller, A.B., C.J. Baines, T. To, and C. Wall. 1992b. "Canadian National Breast Screening Study: 2. Breast Cancer Detection and Death Rates among Women Aged 50 to 59 Years." *Canadian Medical Association Journal* 147 (10): 1477–88.

Miller, Anthony B., Teresa To, Cornelia J. Baines, and Claus Wall. 2000. "Canadian National Breast Screening Study-2: 13-Year Results of a Randomized Trial in Women Aged 50–59 years." *Journal of the National Cancer Institute* 92 (18): 1490–99.

Miller, Anthony B., Teresa To, Cornelia J. Baines, and Claus Wall. 2002. "The Canadian National Breast Screening Study-1: Breast Cancer Mortality after 11 to 16 Years of Follow-Up: A Randomized Screening Trial of Mammography in Women Age 40 to 49 Years." *Annals of Internal Medicine* 137 (5): 305–12.

Miller, Anthony B., Claus Wall, Cornelia J. Baines, Ping Sun, Teresa To, and Steven A. Narod. 2014. "Twenty Five Year Follow-up for Breast Cancer Incidence and Mortality of the Canadian National Breast Screening Study: Randomised Screening Trial." *British Medical Journal* 348.

Moss, Sue M., Howard Cuckle, Andy Evans, Louise Johns, Michael Waller, and Lynda Bobrow. 2006. "Effect of Mammographic Screening from Age 40 Years on Breast Cancer Mortality at 10 Years' Follow-up: A Randomised Controlled Trial." *The Lancet.* 368(9552): 2053–60.

Moss, Sue M., Christopher Wale, Robert Smith, Andrew Evans, Howard Cuckle, and Stephen W. Duffy. 2015. "Effect of Mammographic Screening from Age 40 Years on Breast Cancer mortality in the UK Age Trial at 17 Years' Follow-up: A Randomised Controlled Trial." *The Lancet Oncology* 16 (9): 1123–32.

Mukherjee, Siddhartha. 2017. "Cancer's Invasion Equation." *The New Yorker,* September 4. https://www.newyorker.com/magazine/2017/09/11/cancers-invasion-equation?source=search_google_dsa_paid&gclid=EAIaIQobChMI36qFzvPf7QIVRq-GCh0C1AlFEAAYASAAEgJaAPD_BwE.

Myerson, Rebecca M., Reginald D. Tucker-Seeley, Dana P. Goldman, and Darius N. Lakdawalla. 2020. "Does Medicare Coverage Improve Cancer Detection and Mortality Outcomes?" *Journal of Policy Analysis and Management* 39 (3): 577–604.

National Cancer Institute. 1979. *A History of Cancer Control in the United States, 1946–1971.* Lester Breslow, MD: U.S. National Cancer Institute, Division of Cancer Control and Rehabilitation.

National Center for Health Statistics. 2011. *Health, United States, 2010: With Special Feature on Death and Dying.* Washington, DC: US Government Printing Office.

National Center for Health Statistics. 2019. *Health, United States, 2018.* Hyattsville, Maryland: National Center for Health Statistics.

Nelson, Heidi D., Amy Cantor, Linda Humphrey, Rochelle Fu, Miranda Pappas, Monica Daeges, and Jessica Griffin. 2015. *Screening for Breast Cancer: Systematic Review to Update the 2009 U.S. Preventive Services Task Force Recommendation.* Rockville, MD: Agency for Healthcare Research and Quality.

Nelson, Heidi D., Rochelle Fu, Amy Cantor, Miranda Pappas, Monica Daeges, and Linda Humphrey. 2016. "Effectiveness of Breast Cancer Screening: Systematic Review and Meta-Analysis to Update the 2009 U.S. Preventive Services Task Force Recommendation." *Annals of Internal Medicine* 164 (4): 244–55.

Nyström, Lennarth, Ingvar Andersson, Nils Bjurstam, Jan Frisell, Bo Nordenskjöld, and Lars Erik Rutqvist. 2002. "Long-Term Effects of Mammography Screening: Updated Overview of the Swedish Randomised Trials." *The Lancet* 359 (9310): 909–19.

Nyström, L., S. Wall, L. Rutqvist, A. Lindgren, M. Lindqvist, S. Rydén, J. Andersson et al. 1993. "Breast Cancer Screening with Mammography: Overview of Swedish Randomised Trials." *The Lancet* 341 (8851): 973–78.

Patnick, Julietta, Krys Baker, Isobel Barnes, Valerie Beral, Hongchao Pan, Richard Peto, Gillian Reeves, Mike Richards, and Keith Shaw. 2018. "Nationwide Cluster-Randomised Trial of Extending the NHS Breast Screening Age Range in England: AgeX Trial Protocol. file:///Users/alexandra/Downloads/AgeX%20Protocol%20-%20v5.1%20September%202018%20(MREC%20Approved)-1.pdf.

Pellerin, Renée. 2019. Conspiracy of Hope: The Truth about Breast Cancer Screening. Fredericton, Canada: Goose Lane Editions.

Praga, Claudio, Jonas Bergh, Judith Bliss, Jacques Bonneterre, Bruno Cesana, R. Charles Coombes, Pierre Fargeot et al. 2005. "Risk of Acute Myeloid Leukemia and Myelodysplastic Syndrome in Trials of Adjuvant Epirubicin for Early Breast Cancer: Correlation with Doses of Epirubicin and

Cyclophosphamide." *Journal of Clinical Oncology* 23 (18): 4179–91.

Ravdin, Peter M., Kathleen A. Cronin, Nadia Howlader, Christine D. Berg, Rowan T. Chlebowski, Eric J. Feuer, Brenda K. Edwards, and Donald A. Berry. 2007. "The Decrease in Breast-Cancer Incidence in 2003 in the United States." *The New England Journal of Medicine* 356: 1670–74.

Shapiro, Sam. 1977. "Evidence on Screening for Breast Cancer from a Randomized Trial." *Cancer* 39 (6): 2772–82.

Shapiro, Sam, Philip Strax, and Louis Venet. 1971. "Periodic Breast Cancer Screening in Reducing Mortality from Breast Cancer." *Jama* 215 (11): 1777–85.

Siu, Albert L. 2016. "Screening for Breast Cancer: US Preventive Services Task Force Recommendation Statement." *Annals of Internal Medicine* 164 (4): 279–96.

Sparano, Joseph A., Robert J. Gray, Della F. Makower, Kathleen I. Pritchard, Kathy S. Albain, Daniel F. Hayes, Charles E. Geyer Jr et al. 2018. "Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer." *New England Journal of Medicine* 379: 111–21.

Summary Report of the Working Group to Review the National Cancer Institute-American Cancer Society Breast Cancer Detection Demonstration Projects. 1979. *Journal of the National Cancer Institute* 62 (3): 647–50.

Surveillance, Epidemiology, and End Results Program. 2019. SEER*Stat Database: Mortality–All COD, Aggregated With State, Total U.S. (1969–2017) (Katrina/Rita Population Adjustment), National Cancer Institute, DCCPS, Surveillance Research Program. Underlying mortality data provided by NCHS. (accessed December 4, 2020).

Surveillance, Epidemiology, and End Results Program. 2020. SEER*Stat Database: Incidence–SEER Research Data, 9 Registries, Nov 2019 Sub (1975–2017)–Linked To County Attributes –Time Dependent (1990–2017) Income/Rurality, 1969–2017 Counties, National Cancer Institute, DCCPS, Surveillance Research Program. (accessed December 4, 2020).

Szabo, Liz. 2017. "Doing More Harm than Good? Epidemic of Screening Burdens Nation's Older Patients." *Kaiser Health News*, December 20. https://khn.org/news/doing-more-harm-than-good-epidemic-of-screening-burdens-nations-older-patients/.

Tabár, László, Gunnar Fagerberg, Hsiu-Hsi Chen, Stephen W. Duffy, Charles R. Smart, Adel Gad, and Robert A. Smith. 1995. "Efficacy of Breast Cancer Screening by Age. New Results Swedish Two-County Trial." *Cancer* 75 (10): 2507–17.

Tabár, László, Gunnar Fagerberg, Stephen W. Duffy, and Nicholas E. Day. 1989. "The Swedish Two County Trial of Mammographic Screening for Breast Cancer: Recent Results and Calculation of Benefit." *Journal of Epidemiology & Community Health* 43 (2): 107–14.

USPSTF. 1989. *Guide to Clinical Preventive Services: Report of the US Preventive Services Task Force.* Philadelphia, PA: Williams & Wilkins.

USPSTF. 2002. "Screening for Breast Cancer: Recommendations and Rationale." *Annals of Internal Medicine* 137 (5): 344–46.

USPSTF. 2009. "Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement." *Annals of Internal Medicine* 151 (10): 716–26.

Walter, Louis C., Catherine Eng, and Kenneth E. Covinsky. 2001. "Screening Mammography for Frail Older Women: What Are the Burdens?" *Journal of General Internal Medicine* 16 (11): 779–84.

Welch, H. Gilbert., and William C. Black. 2010. "Overdiagnosis in Cancer." *Journal of the National Cancer Institute* 102 (9): 605–13.

World Health Organization. 2018. *Global Status Report on Road Safety 2018.* Geneva: World Health Organization.

Zackrisson, Sophia, Ingvar Andersson, Lars Janzon, Jonas Manjer, and Jens Peter Garne. 2006. "Rate of Over-Diagnosis of Breast Cancer 15 Years after End of Malmö Mammographic Screening Trial: Follow-up Study." *BMJ* 332: 689–92.

LGBTQ Economics

M.V. Lee Badgett, Christopher S. Carpenter, and Dario Sansone

I n 1995, the first peer-reviewed empirical papers on sexual orientation-based wage differences appeared in economics journals (Badgett 1995a; 1995b). Twenty-five years later, the American Economic Association's Committee on the Status of LGBTQ+ Individuals in the Economics Profession (CSQIEP) made its debut at the 2020 annual meeting of the Allied Social Science Associations. In between, research on lesbian, gay, bisexual, transgender, and queer (LGBTQ) people has grown rapidly. EconLit (the academic literature database maintained by the American Economic Association) indicates that 39 LGBTQ-related articles were published between 1995 and 1999, 162 articles between 2000 and 2009, and 348 between 2010 and 2019. Much of this research has been made possible by the addition of questions to existing surveys that allow LGBTQ respondents to be identified.

The world has also changed over those 25 years. Social movements led by LGBTQ people have grown in influence, leading to the decriminalization of homosexuality in the United States (*Lawrence v. Texas*, 539 US 558 [2003]) and several

■ *M.V. Lee Badgett is Professor of Economics, University of Massachusetts Amherst, Amherst, Massachusetts, and Distinguished Scholar, Williams Institute, University of California at Los Angeles School of Law, Los Angeles, California. Christopher S. Carpenter is E. Bronson Ingram Professor of Economics, Vanderbilt University, Nashville, Tennessee; Research Fellow, National Bureau of Economic Research, Cambridge, Massachusetts; and Research Fellow, IZA (Institute of Labor Economics), Bonn, Germany. Dario Sansone is Lecturer, University of Exeter Business School, Exeter, United Kingdom; and Research Affiliate, IZA (Institute of Labor Economics), Bonn, Germany. Their email addresses are lbadgett@econs.umass.edu, christopher.s.carpenter@vanderbilt.edu, and d.sansone@exeter.ac.uk.*

other countries. Being transgender was depathologized in 2013. Public attention to issues of equality for LGBTQ individuals has resulted in major changes in nondiscrimination policies and the right to marry for same-sex couples in the United States and dozens of other countries (OECD 2020). Research shows that many of these changes have resulted in improved socioeconomic outcomes for some LGBTQ groups. These changes have also generated new research questions for economists and other social scientists.

Our starting point is to look at recent survey data to quantify the number of LGBTQ people and to discuss the quality of these data. We describe the location choices, demographic and economic characteristics, and family structures of the LGBTQ population in the United States compared with cisgender (that is, not transgender) heterosexual individuals. We summarize recent evidence on the socioeconomic effects of legal access to same-sex marriage on a range of family structure, employment, social, and health outcomes. We then turn to a growing body of experimental research that finds evidence of employment discrimination against LGBTQ people. Data on wage gaps are also consistent with a discrimination story for gay and bisexual men; in contrast, lesbian and bisexual women earn more than their heterosexual counterparts. We present some new findings on wage gap trends that show persistent wage gaps for gay/bisexual men but a falling wage advantage for lesbian/bisexual women. We conclude by describing emerging areas of research and new questions in LGBTQ economics.

How Many Individuals Identify as LGBTQ?

At least since the Kinsey Reports on male and female sexual behaviors (Kinsey, Pomeroy, and Martin 1948; Kinsey et al. 1953), social scientists have tried to estimate the percentage of LGBTQ individuals in the population. The task poses challenges with sampling and construction of survey questions, particularly because sexual orientation and gender identity have multiple aspects and because individuals might be reluctant to report a stigmatized characteristic. As one example, we include Q for queer in the LGBTQ umbrella acronym to acknowledge the use of that term by some sexual minorities, although we rarely identify these individuals separately. Similarly, no reliable nationally representative estimates exist for asexual individuals (individuals not sexually attracted to anyone) or intersex individuals (individuals born with sex characteristics that do not fit the typical definitions for male or female bodies). Here, we first describe some prevalence estimates from survey data and then discuss potential challenges regarding their interpretation.

The primary way of counting LGBTQ individuals is to ask in surveys. Table 1 presents a range of survey-based estimates on the proportion of adults who identified as gay, lesbian, bisexual, "something else," and/or transgender. Fourteen high-income countries have included a question on sexual orientation in at least one of their nationally representative surveys and released these data (OECD 2019). On average across these 14 countries, 2.7 percent of adults identified as lesbian, gay, or bisexual,

Table 1
Size of the LGBTQ Population

	OECD (2019)	NHIS		BRFSS
	Selected OECD countries	United States		35 US states
	All adults	All women	All men	All adults
Sexual orientation:				
Straight	[96.2%–98.8%]	95.6%	96.2%	92.1%
Gay/lesbian	[0.5%–1.7%]	1.4%	1.8%	1.7%
Bisexual	[0.3%–2.0%]	1.3%	0.6%	2.2%
Something else	—	0.3%	0.3%	0.7%
Gender identity:				
Transgender	[0.1%–0.3%]	—	—	0.5%
Male to female	—	—	—	0.2%
Female to male	—	—	—	0.2%
Gender nonconforming	—	—	—	0.1%

Source: Tabulations from OECD (2019, p. 17 and p. 19); Authors' tabulations of NHIS 2013–2018, and BRFSS 2014–2018.
Note: This table reports weighted statistics; unweighted statistics are reported in Table B1 in the online Appendix available with this article at the *JEP* website. The shares of individuals who identified as gay, lesbian, bisexual, or something else have been computed in the NHIS and BRFSS data over all (or male/female) respondents (age 18+) who were administered a questionnaire with questions on sexual orientation and gender identity, including those who identified as straight, gay, lesbian, bisexual, something else, who did not know, or who refused to answer. Similarly, the shares of individuals who identified as transgender have also been computed over all respondents who were administered a questionnaire including sexual orientation and gender identity—that is, including those who did not identify as transgender, who did identify as transgender (male to female, female to male, or gender nonconforming), who did not know, or who refused to answer. Data on selected OECD countries have been extracted from the report OECD (2019)

representing 17 million people. In most countries, the share of individuals who identi-fied as gay or lesbian is similar to the share who identified as bisexual.

The next two columns of Table 1 present estimates from two main sources of recent US survey data: the 2013–2018 National Health Interview Surveys (NHIS) and the 2014–2018 Behavioral Risk Factor Surveillance System (BRFSS).[1] The NHIS is nationally representative, while for the BRFSS we observe 35 states in various years that administered a sexual orientation and gender identity module to their core state BRFSS survey and that released their data to the public. As shown in Table 1, 1.4 percent of women (age 18 or more) in the NHIS identified as lesbian, while 1.3 percent identified as bisexual and 0.3 percent identified as "something else" (these are mutually exclusive categories). For men, 1.8 percent identified as gay, while 0.6 percent identified as bisexual and 0.3 percent identified as "something else." The "something else" option was chosen mainly by transgender respondents and by respondents who identified as queer or pansexual, did not use labels to

---

[1] Sections B.1 and D.1 in the online Appendix provide detailed information on the NHIS and BRFSS, respectively. BRFSS data are publicly available on the CDC website (CDC, 2019). BRFSS data are publicly available on the CDC website (CDC 2019).

identify themselves, were in the process of determining their sexuality, or were asexual (Miller and Ryan 2011; Dahlhamer et al. 2014). The BRFSS sample produces similar estimates of the share gay or lesbian but slightly higher estimates of the share bisexual or "something else" compared to the NHIS.

The availability and quality of estimates of the transgender population are limited. For example, only three OECD countries have asked questions about gender identity on a nationally representative survey (OECD 2019). The percentage in the adult population who identify as transgender ranges from 0.1 percent in Chile to 0.3 percent in the United States (OECD 2019). Table 1 shows that the BRFSS data for 35 states indicate that 0.5 percent of adults identified as transgender. Of these, 0.2 percent identify as transgender male-to-female, 0.2 percent identify as transgender female-to-male, and 0.1 percent identify as transgender gender nonconforming.

Data Quality on LGBTQ People

How much should we trust self-reported information on sexual orientation? We know that survey results vary depending on how researchers classify sexual minorities (Laumann et al. 2000; Gates 2011). For example, survey questions can ask about romantic, sexual, and/or emotional attraction to members of the same sex; about actual same-sex sexual behavior; or about whether an individual self-identifies as gay, lesbian, bisexual, or queer. Of these, the share of people reporting same-sex attraction is consistently larger than the share reporting same-sex sexual behavior or the share identifying as LGBTQ. How questions are worded, and the degree of privacy and anonymity afforded to survey respondents—say, in-person interviews versus computer-assisted ones—can also affect these estimates (Robertson et al. 2018). In the United States, for example, the average estimate of self-identified lesbian, gay, or bisexual people across surveys is 3.8 percent, but this total increases to 4.6 percent when focusing only on self-administered modules (OECD 2019). Coffman, Coffman, and Ericson (2017) used "list experiments" designed to elicit the true rate of non-heterosexuality in a non-representative Amazon Mechanical Turk sample (which is younger, more highly educated, and more liberal than the general population). The authors found that rates of non-heterosexual identity implied by the list experiment were 65 percent larger than rates based on direct self-reports.[2]

Research on sexual orientation and gender identity survey questions continues to expand, but no consensus yet exists on best practices for gender identity. The questions about sexual orientation on large federal surveys such as the NHIS have undergone extensive testing to understand why respondents choose particular

---

[2] Researchers have also used Internet data which affords users anonymity to suggest that same-sex attraction and behavior are higher than reported in surveys (Pornhub Insights 2017; Stephens-Davidowitz 2017).

answers and why some response options pose challenges.[3] A working group of statisticians from federal statistical agencies reviewed evaluations of sexual orientation and gender identity questions and found that existing measures of sexual identity have performed well (Federal Interagency Working Group 2016).

Of course, one may worry that willingness to self-identify as LGBTQ on surveys is correlated with unobserved characteristics such as family and social support in a way that could bias estimates of outcomes under study. Given that attitudes toward sexual minorities have improved over the recent past, one approach is to examine whether rates of reporting LGBTQ status have increased over time. Unfortunately, most surveys with direct questions on sexual orientation or gender identity have not included them over a sufficiently long period of time as to be informative in this context. But there are exceptions. First, the Gallup corporation found that the share of Americans who identify as LGBT increased from 3.5 percent in 2012 to 4.5 percent in 2017 (Newport 2018). But only the youngest cohort—born between 1980 and 1999—showed any increase in prevalence, rising from 5.8 percent in 2012 to 8.2 percent in 2017, and much of that change was accounted for by women. Second, a more detailed analysis of three waves of data from the 2002–2013 National Surveys of Family Growth found an upward time trend in the percentage of women age 18–45 reporting a bisexual identity and the percentage reporting a same-sex sex partner (England et al. 2016). It is not possible to distinguish whether those trends reflect an increased willingness of some subgroups of LGBTQ people to report their status or whether the trends capture changes in sexual behavior and identity for those subgroups.

Fortunately, another dataset allows us to indirectly assess trends in reporting sexual minority status. Specifically, a large body of prior research pioneered by Black et al. (2000) uses the American Community Survey to look at individuals in same-sex couples who are very likely to be sexual minorities in cohabiting same-sex romantic relationships. The American Community Survey identifies a primary reference person, defined as "the person living or staying here in whose name this house or apartment is owned, being bought, or rented." The American Community Survey also collects information on the relationship to the primary reference person for all individuals living at the same address, and the range of possible relationships includes husband, wife, and unmarried partner (as a different category than roommate or other nonrelative).[4] We can thus identify same-sex couples by

---

[3]Some measurement problems come from those non-LGBTQ people for whom sexual identity is not a particularly salient self-concept. Some do not understand the term "heterosexual" and refuse to answer or respond with "something else" or "I don't know" (Ridolfo, Miller, and Maitland 2012). This finding influenced the design of the sexual orientation question on the National Health Interview Survey, for example, which uses "straight, that is, not gay" instead of "heterosexual" in the English language survey (Dahlhamer et al. 2014).

[4]Section C.1 in the online Appendix provide detailed information on the American Community Survey and discuss possible measurement error issues highlighted in the literature (like misclassification of different-sex couples as same-sex couples) and how they have been addressed. To reduce measurement error, the relationship categories of spouse and unmarried partner have recently been changed on the Census, American Community Survey, and Current Population Survey to include opposite-sex husband/

linking the household primary reference person with their same-sex spouse or unmarried partner, and we can also see households with two people of the same sex who do not report being in a same-sex partnership. As a result, we can classify thousands of couples as same-sex in each year of the survey data. Our intuition here is that increasing trends over time in the share of households with two same-sex adults that are same-sex couples would be consistent with increased reporting of same-sex couple status (and, by implication, minority sexual orientation) over time.

Thus, we compare two measures: 1) the share of all two-adult households composed of two same-sex adults; and 2) the share of all households with two same-sex adults that is composed of same-sex couples (either same-sex unmarried partners or married same-sex spouses).[5] Changes in the first series could plausibly reflect changes in coupling and cohabitation decisions of sexual minorities, while changes in the second series are more likely to reflect changes in reporting decisions among sexual minorities conditional on same-sex cohabitation. The share of all two-adult households composed of two same-sex adults increased modestly over the time period (by about 10 percent) while the share of households with two same-sex adults that was composed of people who describe themselves as same-sex couples exhibited closer to a 50 percent increase. Taken together these patterns point to relatively larger increases in *reporting* of same-sex relationships conditional on same-sex cohabitation over the past decade, with a more modest change in same-sex cohabitation itself. Figure C1 in the online Appendix shows the trends from 2008 to 2018.

If these changes over time are also reflected in changes in observed demographic characteristics, how should we think about their effects on estimates of gaps in economic outcomes between sexual minorities and heterosexuals? We look at trends in the share of all couples (both different-sex and same-sex) that are same-sex couples across age, race, education, and region. Several patterns emerge. First, the same-sex couple shares increased for every demographic group we examine.[6] Second, the increase in the same-sex couple share was slightly higher for younger people (18–45) relative to older people (46+). Third, the increase in the same-sex couple share was notably higher for Black individuals relative to other race groups. Fourth, the increase in the same-sex couple share was slightly smaller for the less educated, relative to the highly educated. Finally, there were no noticeable differences in the trends over time in the same-sex couple shares across regions. Overall, these patterns are consistent with younger, slightly more educated, and more racially diverse sexual minorities being more represented among the sample of same-sex couples in recent years.

wife/spouse, same-sex husband/wife/spouse, opposite-sex unmarried partner, and same-sex unmarried partner as separate categories. ACS data are publicly available through IPUMS-USA at the University of Minnesota (Ruggles et al. 2020).

[5] These trends are reported in Panels B and C of online Appendix Figure C1, respectively. For the sake of completeness, Panel A reports the trend in the share of all households composed of exactly two adults.

[6] For details of the calculations, see Figure C2 in the online Appendix. These findings are also in line with the general trend plotted in Panel D of Figure C1 and show the share of all couples (both different-sex and same-sex) that are same-sex couples without splitting the sample based on demographic characteristics.

Taken as a whole, the evidence suggests that sexual orientation and gender identity questions perform reasonably well, even though not all LGBTQ people are willing to report as such on surveys. Also, the analysis of the American Community Survey supports the idea that sexual minorities are becoming more willing to disclose their status on surveys over time. However, it is important to note that willingness to report being LGBTQ on a survey does not necessarily imply an individual's openness in other settings being studied, such as the workplace. If the willingness to report on a survey is related to openness about being LGBTQ in the setting being studied, then the economic outcomes observed by sexual orientation or gender identity might be biased. More research is needed to understand which LGBTQ people reveal their identities in various data, perhaps linking survey data to administrative records. But in the research review below, the consistency of many findings over time, across datasets and research designs, and across measures suggest that reporting bias is not likely to be the source of the broad findings observed.

Location Choices of Sexual Minorities

Historically, income, fertility, and attitudes in the general population have affected the location and migration choices of sexual minorities, an early topic studied by economists (Black et al. 2002). The American Community Survey also provides very large sample sizes of same-sex couples, which allows for meaningful comparisons across geographic areas.

Where were same-sex couples located at the time of the survey? Washington, DC, had by far the highest number of individuals in same-sex couples as a percent of all individuals in couples (6.8 percent), followed by Vermont (1.89 percent), Massachusetts (1.80 percent), and Rhode Island (1.61 percent).[7] The states with the lowest same-sex couple shares were Wyoming (0.57 percent), North Dakota (0.59 percent), and South Dakota (0.63 percent). If we focus on metropolitan areas, the metropolitan areas with the largest same-sex couple shares were Ithaca, NY; San Francisco, CA; Santa Fe, NM; and Santa Rosa, CA.[8]

Individuals in same-sex couples were more likely than individuals in different-sex couples to reside in a different state from their birth: for men, 45 percent of those in same-sex couples lived in the state in which they were born, compared to 59 percent of those in different-sex couples; for women, 52 percent of those in same-sex couples lived in the state in which they were born, compared with 60 percent of

---

[7] Figure C3 in the online Appendix plots the weighted and unweighted distribution of same-sex couples across states as well as the states with the largest number of same-sex couples. Weighted percentages are reported here.

[8] Table C3 in the online Appendix reports weighted and unweighted percentages as well as the metropolitan areas with the largest number of same-sex couples. In addition, Tables C4-C5 report the top 20 metropolitan areas separately for women and men in same-sex couples, while Tables C6-C8 report the top 20 cities in terms of number and share of individuals in same-sex couples, jointly and separately by sex.

those in different-sex couples. These gaps do seem to be closing, albeit for women more than for men, consistent with improving attitudes reducing the need for sexual minorities to migrate to more LGBTQ-friendly environments.[9]

Today, the concentration of LGBTQ people in cities and states with progressive policies could reflect an influence of the social climate on reporting LGBTQ status on surveys (National Academies of Sciences, Engineering, and Medicine 2020), as noted earlier. In addition, research in political science suggests that the size and strength of the LGBTQ population is positively correlated with the passage of sexual orientation and gender identity nondiscrimination laws (Taylor, Haider-Markel, and Rogers 2019). However, it is important to emphasize that LGBTQ people do not live exclusively in cities or on the coasts: for example, between 2.9 and 3.8 million LGBTQ people are estimated to live in rural areas (MAP 2019).

Demographic and Economic Characteristics of People Who Identify as LGBTQ

In this section, we describe the demographic profiles of people who identify as LGBTQ, and we compare these with the characteristics of heterosexual and cisgender people. We use data from the 2013–2018 National Health Interview Surveys, the 2008–2018 American Community Surveys, and the 2014–2018 waves of the Behavioral Risk Factor Surveillance System.
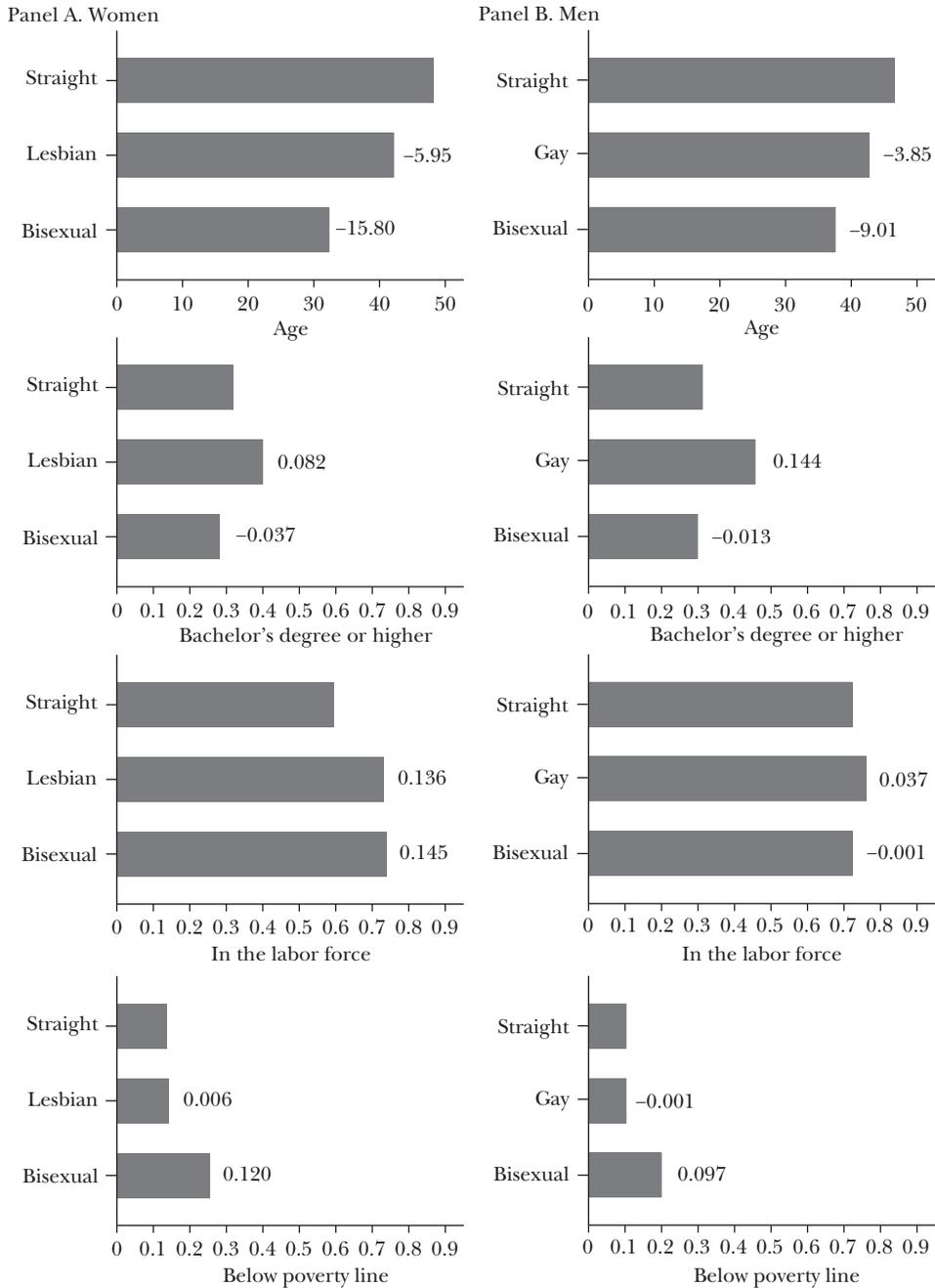
Differences by Sexual Orientation

The National Health Interview Survey is the only nationally representative and publicly available survey in the United States with direct information on sexual orientation that produces large sample sizes of sexual minorities and has high quality data on demographic and economic characteristics. Here, we focus on dimensions where there are substantial differences related to sexual orientation.

The top two panels of Figure 1 show that straight men and women were older than those who identified as lesbian and gay, who in turn were older than those who identified as bisexual. This pattern fits with the hypotheses that norms about sexual conduct are evolving or that younger people are more willing to volunteer a non-heterosexual identity on survey questions. The second row of bar graphs in Figure 1 shows a measure of education: the share of each group that reported having a bachelor's degree or higher. Those who identified as gay or lesbian had notably higher education than those who identified as either straight or bisexual.

---

[9]While the share of women residing in their birth state did not change from 2008–2018 for women age 46 or older in different-sex couples, younger women (age 18–45) in different-sex couples, or older women in same-sex couples, it slightly increased for younger women in same-sex couples. The pattern for men, however, is weaker. For details, see Figure C4 in the online Appendix.

Figure 1
Demographic Characteristics by Individual Self-Reported Sexual Orientation

Note: Weighted statistics. Sample includes all sample adults (age 18+). Tables B3 and B4 in the online Appendix reports a more extensive set of weighted summary statistics by sexual orientation, including individuals who answered "don't know" to the sexual orientation question, who refused to answer this question, or who skipped the question. Tables B5 and B6 report unweighted summary statistics. All variables are described in Section A of the online Appendix. Numbers next to bars report mean differences with respect to straight individuals.

The third row of bar graphs in Figure 1 shows labor market participation. Here, straight women had lower labor force participation than lesbian and bisexual women, while gay men were slightly more likely to be in the labor force than straight or bisexual men. Finally, the fourth row of bar graphs shows the share with family incomes below the poverty line. Bisexual women and bisexual men had the highest share below the poverty line in each group. Such findings are in line with previous studies (Badgett 2018) and consistent with the fact that bisexual individuals may experience greater stigma, less social acceptance, and less community support than lesbian women or gay men (Herek et al. 2010). High rates of poverty are also observed for individuals who identified as "something else" when asked about their sexual orientation (Tables B3 and B4 in the online Appendix).

The NHIS data does not show especially large differences in race or ethnicity across most of these groups. We also do not see big differences in having health insurance or in having a disability across most of these groups, but previous research suggests some disparities among heterosexual and non-heterosexual individuals in terms of health insurance and access to health care (Ash and Badgett 2006; Gonzales and Blewett 2014) and health outcomes (Institute of Medicine 2011).

Unsurprisingly, the main distinctions between different-sex and same-sex couples in the 2008–2018 American Community Survey data are much the same as in the NHIS data on sexual orientation.[10] For example, women and men in same-sex couples were on average a few years younger than individuals in different-sex couples. Individuals in same-sex couples were twice as likely to be in college at the time of the American Community Survey than those in different-sex couples. Adults in same-sex couples were more likely to accumulate higher human capital: 43 percent of women in same-sex couples and almost half of men in same-sex couples had a bachelor's degree or a higher educational level, compared to 34 percent of individuals in different-sex couples (Black, Sanders, and Taylor 2007; Sansone and Carpenter 2020).[11] Interestingly, the most striking gaps are at the highest level: both women and men in same-sex couples had a higher likelihood than those in different-sex couples to have received a master's degree, an advanced professional degree, or a doctoral degree (21 percent versus 13 percent for women; 21 percent versus 14 percent for men).

Again, we observe higher labor force participation rates for individuals in same-sex couples than for individuals in different-sex couples. The gap between women in same-sex and different-sex couples is particularly striking (76 percent versus 62 percent), as is the gap in proportions of women working full-time (58 percent versus 41 percent). These gaps are partially explained by different fertility levels and by the fact that women in same-sex couples are more likely to be primary earners than women in different-sex couples (Antecol and Steinberger 2013). The data

[10] All the descriptive statistics discussed in this section have been computed using IPUMS person weights and presented in Table C9 in the online Appendix. Table C10 reports unweighted summary statistics.

[11] However, studies that focus on a single cohort, self-reported LGBTQ individuals, or on young adults found more complicated patterns of education, not always supporting the conclusion that LGBTQ individuals have higher educational levels than their heterosexual counterparts (Pearson and Wilkinson 2017; Sansone 2019b).

from the American Community Survey also suggest that women in same-sex couples are penalized from a double gender pay gap. Although labor force participation for women in same-sex couples was high, their personal income was lower on average than that of men in same-sex couples, and their family income was the lowest among all couple types.

Differences by Transgender Status

The National Health Interview Survey and the American Community Survey do not include information on gender identity. Thus, Table 2 reports summary statistics by gender identity using data from the Behavioral Risk Factor Surveillance System. A transgender identity does not imply a particular sexual orientation: a majority of transgender respondents (61 percent) identified as straight. The comparisons here are between cisgender women, cisgender men, male-to-female transgender individuals, female-to-male transgender individuals, and individuals who are gender nonconforming.

In line with Herman et al. (2017), transgender individuals, especially those who are gender-nonconforming, were substantially more likely to be young (age 18–34) than cisgender respondents. They were also more likely to be Black or Hispanic. In contrast with the higher educational levels previously reported among gay and lesbian adults, a smaller percentage of transgender individuals graduated from college, although a larger share of transgender individuals than cisgender individuals were in school at the time of the survey.

While health disparities among transgender and cisgender individuals have been analyzed (National Academies of Sciences Engineering and Medicine 2020), it is worth emphasizing that transgender individuals were less likely to have health insurance coverage and had considerably higher disability rates. In terms of employment outcomes, labor force participation rates among transgender individuals were between those of cisgender women and cisgender men (in line with the findings of Carpenter, Eppink, and Gonzales 2020). Nevertheless, their unemployment rates were much higher: for instance, male-to-female transgender respondents had double the unemployment rate of cisgender men. Relatedly, transgender respondents were more likely to live in households with low incomes.

Relative to transgender individuals who describe themselves as male-to-female or female-to-male, transgender individuals who describe themselves as gender-nonconforming were the youngest; the least likely to be white or with a college education; and the most likely to be Hispanic, without health insurance coverage, with a disability, not in the labor force, unemployed, or in a low-income household.

LGBTQ Families

Next, we describe the family structures of LGBTQ adults, paying particular attention to marriage, partnership, and cohabitation as well as the presence of children. In some ways, the patterns we find are updates of previously published

Table 2

Demographic Characteristics by Gender Identity

	Cisgender women	Cisgender men	Transgender Male to female	Transgender Female to male	Transgender Non-conforming
Demographic characteristics:					
Age 18–34	0.26	0.29	0.33	0.39	0.54
Age 35–49	0.23	0.24	0.21	0.24	0.14
Age 50–64	0.27	0.27	0.29	0.20	0.19
Age 65+	0.23	0.19	0.17	0.17	0.12
White	0.77	0.76	0.69	0.69	0.65
Black	0.14	0.12	0.16	0.20	0.16
Asian	0.06	0.06	0.08	0.06	0.09
Other races	0.04	0.05	0.06	0.05	0.10
Hispanic	0.15	0.15	0.19	0.19	0.22
Education and health:					
College education	0.27	0.26	0.13	0.12	0.18
Student	0.06	0.05	0.06	0.09	0.13
Any health insurance coverage	0.90	0.87	0.81	0.80	0.85
Any disability	0.25	0.20	0.33	0.37	0.43
Employment and income:					
In the labor force	0.55	0.70	0.63	0.58	0.55
Unemployed (versus employed)	0.09	0.08	0.16	0.11	0.20
Total household income below $50,000	0.54	0.47	0.66	0.66	0.63
Total household income below $15,000	0.12	0.08	0.20	0.19	0.16
Observations	538,135	408,503	1,868	1,330	877

Source: BRFSS 2014–2018

Note: Weighted statistics. "Observations" refers to the total number of respondents (age 18+) in the relevant sub-group. In addition to the statistics presented in this table, Tables D2-D3 in the online Appendix report summary statistics by sexual orientation for men and women, respectively: these tables show gaps between heterosexual, gay, lesbian, bisexual, and queer individual qualitatively similar to those highlighted in Figure 1 using NHIS data. Table D4 in reports unweighted summary statistics by gender identity, while Table D5 reports weighted summary statistics also for individuals who answered "don't know" to the gender identity question, who refused to answer this question, or who skipped the question.

work (Black et al. 2000; Black, Sanders, and Taylor 2007; Carpenter and Gates 2008; Oreffice 2011; Gates 2015a; 2015b; Sansone 2019a), though we make two new contributions. First, we describe family structures of LGBTQ individuals in addition to same-sex couples. Second, we provide the most up-to-date, nationally representative estimates of partnership and marriage rates for LGBTQ-identified individuals.

Marital Status and Presence of Children

Using data from the 2013–2018 National Health Interviews Survey, Table 3 shows that partnership rates—the share of individuals who reported being married or living with a partner—were similar for lesbians (53 percent) and heterosexual women (59 percent) but were lower for bisexual women (40 percent). While the

Table 3
Marital Status by Sexual Orientation

	Women			Men		
	Straight	*Lesbian*	*Bisexual*	*Straight*	*Gay*	*Bisexual*
Married	0.52	0.27	0.22	0.57	0.21	0.20
In an unmarried couple	0.07	0.26	0.18	0.07	0.22	0.10
Divorced	0.11	0.06	0.08	0.08	0.05	0.10
Separated	0.02	0.01	0.04	0.02	0.01	0.02
Widowed	0.09	0.01	0.02	0.03	0.01	0.02
Never married	0.20	0.38	0.47	0.24	0.51	0.56
Observations	97,909	1,424	1,235	80,191	1,752	509

Source: NHIS 2014–2018
Note: Weighted statistics. Sample includes all sample adults (age 18+). Individuals are coded as children based on their relationship with the household primary reference person. Online Appendix Table B7 reports statistics for individuals who identified as "something else" when asked about their sexual orientation. Table B8 contains unweighted statistics.

overwhelming majority (88 percent) of partnered heterosexual women reported being married as opposed to being in an unmarried couple, for lesbian women partnership was nearly equally split between the two categories.

For men in Table 3, we observe that heterosexual men had much higher partnership rates (64 percent) than gay men (43 percent) or bisexual men (30 percent). These partnership rates using recent nationally representative samples for gay men and lesbians are very similar to those reported by Carpenter and Gates (2008) for adults in California. Also, among partnered heterosexual men, the overwhelming majority (89 percent) reported being married as opposed to being in an unmarried couple, while for gay men partnership is nearly equally split between the two categories. All sexual minority women were much more likely than heterosexual women to report that they were never married, and the identical pattern is observed for men.

Figure 2 presents the share of individuals with children under age 18 and children under age five in the household separately by sexual orientation.[12] Unsurprisingly, Panel A shows that heterosexual women were the most likely to have children in the household, though the associated share of bisexual women with children was nearly identical (33 versus 32 percent). Fully 20 percent of self-identified lesbian women had children in the household, with 7 percent having young children. For men in Panel B of Figure 2, while the share of heterosexual men with children in the household is very similar to the associated share of

---

[12] The belief that children raised by LGBTQ parents might be harmed in some way has been extensively studied by social scientists for decades without finding evidence of developmental or other harms (for example, Patterson 2005; Boertien and Bernardi 2019).

Figure 2
Presence of Children by Sexual Orientation

Panel A. Women

Panel B. Men
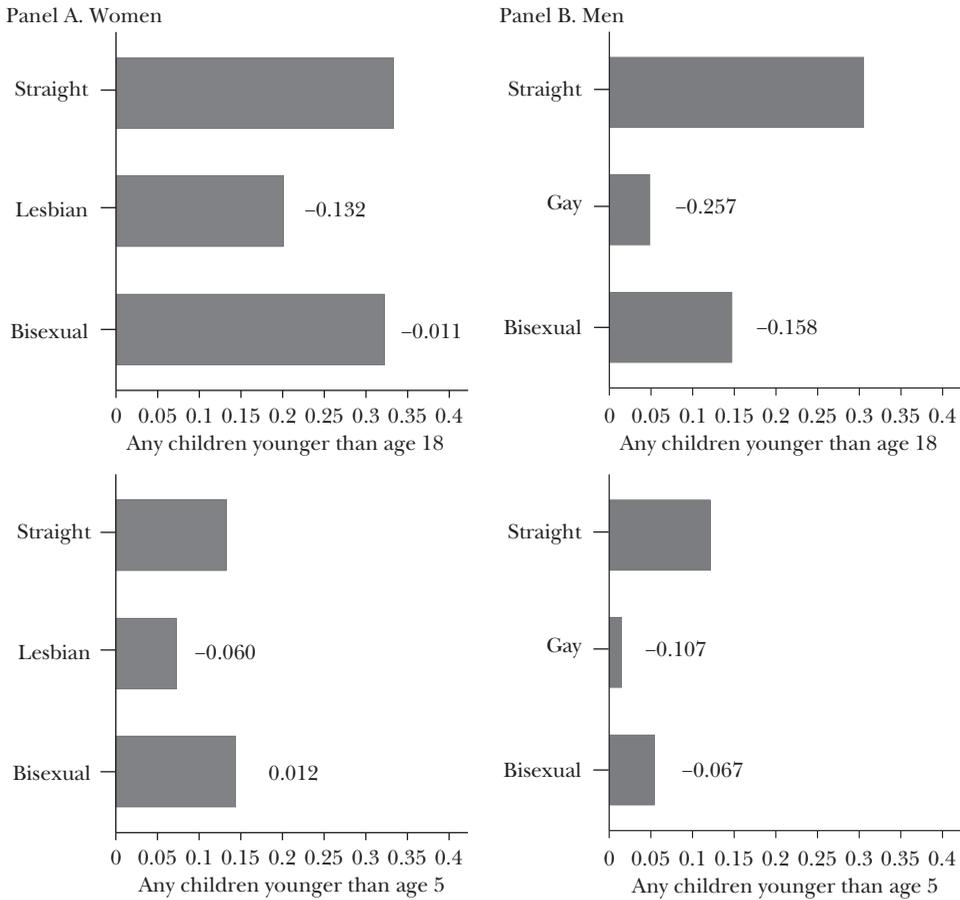
Source: NHIS 2014–2018.
Note: Weighted statistics. Individuals are coded as children based on their relationship with the household primary reference person. Only sample adults (age 18+) who are the household primary reference person, or their spouse, or their unmarried partners have been considered. All statistics are reported in tabular form in Table B7 in the Online Appendix. Table B8 reports unweighted summary statistics. Numbers next to bars report mean differences with respect to straight individuals.

heterosexual women (as expected), far fewer gay and bisexual men had children in the household.

Based on the NHIS data, we cannot say whether the child was from a previous relationship, including previous different-sex marriages and partnerships. Biological differences in the costs of adding children to a household could contribute substantially to the childrearing differences across groups. A complicating factor

here is that the vast majority of *both* partnered bisexual women (87 percent) and partnered bisexual men (90 percent) were in different-sex relationships (as opposed to same-sex relationships). The lower rates of partnership for gay and bisexual men compared to those for lesbian and bisexual women could also play a role in the lower rates of childrearing, as could discrimination against male same-sex couples by foster care or adoption agencies (Mackenzie-Liu, Schwegman, and Lopoo forthcoming).

These differences in rates of children present in the household are also present in comparisons of same-sex couples and different-sex couples using the American Community Survey (see Tables C11 and C12 in the online Appendix). Using detailed information on relationship of children to the household primary reference person in the American Community Survey, we found that men and women in same-sex couples were much more likely to have adopted children or stepchildren in the household than men and women in different-sex couples (as also found earlier by Gates 2015a). The overwhelming majority (95 percent) of different-sex households with any children were composed of households with biologically-related children.

Partnership Dynamics

We can also use the NHIS data to explore partnership dynamics.[13] Although several studies have examined homogamy and assortative mating among same-sex couples (Jepsen and Jepsen 2002; Ciscato, Galichon, and Goussé 2020), far less work has examined partnership formation among all sexual minority individuals, including single gay men, lesbians, and bisexual individuals. Specifically, we can compare those not in a partnership to those in a partnership, and we then further divide the partnered group into those living with a same-sex partner and those who are married. Because the data consists of cross-sectional snapshots, comparisons between non-partnered and partnered individuals could reflect any potential effects of partnership on some of the outcomes (for example, investments in education), and these will be empirically indistinguishable from selection effects.

Compared to non-partnered lesbian women, partnered lesbian women were older (average age of 44.3 versus 39.3 years), more likely to be white (84 versus 70 percent), more likely to have children in the household (25 versus 10 percent), more likely to have a bachelor's degree (52 versus 27 percent), less likely to have a disability (18 versus 24 percent), more likely to be in the labor force (80 versus 64 percent), less likely to be poor (7 versus 23 percent), and more likely to own a home (63 versus 45 percent). There are complex associations related to marriage: on the one hand, compared to lesbians in a same-sex unmarried partnership, married lesbians were older, more likely to have children in the household, more likely to have a college education, and more likely to have health insurance than lesbians who were in a same-sex unmarried partnership. These patterns are

---

[13] Tables B9–B11 in the online Appendix present detailed cross-tabulations of demographic characteristics by partnership status and sexual orientation.

similar to the patterns observed between non-partnered and partnered lesbians. On the other hand, compared with lesbians in a same-sex unmarried partnership, married lesbians were slightly less likely to be in the labor force, less likely to be full-time employed, and—despite these patterns—were much less likely to have low family incomes and to be in poverty compared with lesbians who were in a same-sex unmarried partnership. Married lesbians were also much more likely to be homeowners than lesbians who were in a same-sex unmarried partnership. One hypothesis for explaining these patterns is that married lesbians may be more likely to have a household division of labor in which one person is the predominant income-earner.

For gay men, the patterns regarding non-partnered and partnered individuals are broadly similar to those observed for lesbian women. Among those gay men who are partnered, the gay men who reported being married were older (average age of 48 years versus 43.2 years) and less likely to be working full-time (60 versus 66 percent) than gay men who reported living with a same-sex partner. However, for gay men the association of marriage (conditional on same-sex partnership) with race, US citizenship, college education, disability, personal and family income, poverty status, and homeownership was either more muted or went in the opposite direction than for lesbians.

Family Outcomes among Transgender Individuals

For patterns on family outcomes by self-identified transgender status, we turn again to data from the 2014–2018 Behavioral Risk Factor Surveillance System. As shown in Table 4, transgender people were less likely to be married than cisgender people, though between 33 and 41 percent of transgender individuals reported being married, and a substantial share (3–11 percent) of the transgender population also reported being in an unmarried couple. Many transgender individuals resided with children: 41 percent of individuals who described themselves as transgender, female-to-male were living in households with children present, which is actually higher than the associated shares for cisgender men and women.

It is worth mentioning that the Behavioral Risk Factor Surveillance System questionnaire does not specifically ask about the number of respondents' children, but only the number of children living in the household. The BRFSS shares of transgender people with children in the household is consistent with previous studies and other surveys from the United States (Meyer et al. 2017), although it is higher than estimates from Dutch administrative registry data (Geijtenbeek and Plug 2018).

Legal Access to Same-Sex Marriage

Prior to 2004, LGBTQ people in same-sex couples in some US states had access to civil unions and/or domestic partnerships. These policies varied greatly; some states granted nearly all the same benefits as were afforded to different-sex married couples, while in other states that status was primarily symbolic. Massachusetts was

Table 4
Family Outcomes by Gender Identity

	Cisgender women	Cisgender men	Transgender, male to female	Transgender, female to male	Transgender, gender non-conforming
Marital status:					
Married	0.50	0.53	0.41	0.39	0.33
In an unmarried couple	0.04	0.05	0.06	0.03	0.11
Divorced	0.12	0.10	0.12	0.10	0.08
Separated	0.03	0.02	0.04	0.03	0.03
Widowed	0.11	0.04	0.05	0.09	0.07
Never married	0.21	0.27	0.32	0.36	0.39
Presence of children:					
Any child in the household	0.38	0.34	0.28	0.41	0.28
Observations	538,135	408,503	1,868	1,330	877

Source: BRFSS 2014–2018
Note: Weighted statistics. "Any child in the household" only includes children less than 18 years of age. "Observations" refers to the total number of respondents (age 18+) in the relevant sub-group. Table D6 in the online Appendix contains unweighted statistics.

the first state to legalize same-sex marriage in the United States in 2004 through a state Supreme Judicial Court ruling.

The US Supreme Court considered same-sex marriage in both *US v. Windsor* (570 US 744 [2013]) and *Obergefell v. Hodges* (576 US 644 [2015]). In *Windsor,* the Supreme Court ruled that same-sex marriages legally performed in states that permitted it were entitled to be recognized as full legal marriages by the federal government, striking down the part of the 1996 Defense of Marriage Act that had defined marriage as between one man and one woman for federal purposes. Two years later, the Supreme Court legalized same-sex marriage nationwide in *Obergefell,* effectively ruling that denying same-sex couples the right to legal marriage violated the Due Process Clause and Equal Protection Clause of the Constitution.

Economists contributed greatly to the same-sex marriage debate, both as expert witnesses in the court cases (Badgett 2009) and as social scientists evaluating the effects of legal access to same-sex marriage.[14] For example, a sizable literature has evaluated the effects of changes in legal recognition of same-sex couples on marriage take-up using difference-in-differences methods. Dillender (2014) found

---

[14]Parallel legal changes in recognition of same-sex relationships also occurred in countries throughout North and South America, Western Europe, and Australia. Economists have, in particular, studied how the diffusion of these policies within and across Europe affected outcomes such as different-sex marriage rates (Trandafir 2015) and sexually transmitted infections (Dee 2008). A literature in economics and political science has found that legal recognition of same-sex relationships improved attitudes toward sexual minorities in both the United States (Flores and Barclay 2016) and Europe (Aksoy et al. 2020), though Ofosu et al. (2019) found evidence of backlash effects where the judicial imposition of legal same-sex marriage led to increased implicit and explicit bias against sexual minorities.

no evidence that granting relationship recognition to same-sex couples changed different-sex US marriage rates. Carpenter (2020) studied legal same-sex marriage in Massachusetts in 2004 and found significantly increased marriage take-up among self-identified gay men, lesbians, and bisexual women, with no negative effect on marriage among heterosexuals. Carpenter et al. (forthcoming) examined data from 2000–2018 spanning the rollout of legal access to same-sex marriage and found that marriage equality in an individual's state significantly increased take-up of marriage among individuals in same-sex households, with no effects for individuals in different-sex households.

Regarding economic and social outcomes, economists have studied the effects of legal same-sex marriage using timing variation in a difference-in-differences framework. Sansone (2019a), using variation across states in the timing of legal access to same-sex marriage and data from the American Community Survey, the Survey of Income and Program Participation, and Google Trends, found evidence that marriage equality significantly increased probabilities of being employed among individuals in same-sex couples. He argued that a key mechanism was reduction in discrimination against sexual minorities. Hansen, Martell, and Roncolato (2019) examined data from the American Community Survey, March Current Population Survey, and American Time Use Surveys and found that legal access to same-sex marriage did not change labor supply decisions of men in same-sex couples, but significantly reduced hours of work for women in same-sex couples, particularly for the lower earner within the household. They suggested that marriage equality increased specialization within the households of female same-sex couples. Hamermesh and Delhommer (forthcoming) find that legal same-sex marriage induced greater investments into same-sex relationships: specifically, marital surplus and homeownership for same-sex couples was not strongly related to the duration of the relationship when there was no legal same-sex marriage; in contrast, legal access to same-sex marriage was associated with positive returns to relationship duration with respect to marital surplus and homeownership. Miller and Park (2018) find that legal same-sex marriage was associated with significant increases in applications for mortgage credit for same-sex couples. Carpenter et al. (forthcoming) find that legal access to same-sex marriage increased health insurance coverage and access to care for men in same-sex households. As we note below, more research is needed to understand the full range of outcomes that were affected by same-sex marriage legalization.

Labor Market Discrimination and the LGBTQ Population

Historically, LGBTQ people have faced discrimination in employment, with outright bans in federal employment and, in some places, teaching professions in the 1950s and 1960s, continuing to the recently lifted ban on transgender people serving in the military. It is plausible that employment discrimination still exists to some degree. However, such discrimination is now illegal as a result of the 2020

Supreme Court ruling in *Bostock v. Clayton County* in which the Court found that discrimination against LGBTQ people is a form of sex discrimination prohibited under Title VII of the Civil Rights Act of 1964. This ruling affirmed the Equal Employment Opportunity Commission's 2013 decision to treat sexual orientation and gender identity complaints as sex discrimination charges. These two actions expanded discrimination protections beyond the 22 states that outlawed sexual orientation and/or gender identity discrimination in employment to the entire country.

Researchers have approached the question of discrimination against LGBTQ people with methods similar to studies of discrimination with respect to race, sex, disability, and other characteristics. In this section, we review experimental studies of the job application process, observational studies of wage differences by LGBTQ status, other evidence of discrimination, and the role of public policy. We also assess potential theoretical explanations for observed employment and wage outcomes.

Experimental Evidence of Discrimination

Studies of employment discrimination in many areas have increasingly used experimental methods to assess discrimination in job searches (Neumark 2018), and a growing body of research using similar methods examines discrimination against LGBTQ people (Valfort 2017; Neumark 2018; Granberg, Andersson, and Ahmed 2020). These studies provide the strongest evidence that being LGBTQ causes differential treatment.

In these studies, sexual orientation is usually indicated on a resume through membership in an LGBTQ organization or on a social media profile—for example, a man indicating that he is interested in men (Acquisti and Fong 2020). Studies sometimes incorporate common local features of job applications to enhance variation in other potentially relevant characteristics, such as including pictures of women wearing flowing clothes (a more feminine gender presentation) and women wearing more tailored clothes (indicating a more masculine gender presentation, Weichselbaumer 2003). Gender identity is usually indicated by organization membership, name markers, or gender/sex markers.

The large majority of such studies found that LGBTQ job candidates were significantly less likely to be invited for an interview or to be offered a job. The experimental studies also suggest some of the underlying reasons for the discrimination observed. Employers may be displaying their distaste for employing LGBTQ people in their differential treatment of effectively identical applicants. Some studies have instead tested for evidence of statistical discrimination, which could be at work if employers use applicants' LGBTQ status to infer job-related characteristics (Arrow 1973). Studies testing for statistical discrimination hypothesize that employers assume that LGBTQ people are gender nonconforming and less likely to adhere to behavior that reflects a "legitimate" job requirement (which may be debatable in many contexts). Tilcsik (2011) found that employers discriminate more against gay male applicants when job ads seek stereotypical male

qualities like aggressiveness, decisiveness, assertiveness, and ambition. Lesbians, in contrast, should have an advantage in seeking work because they are less likely to have children. The experimental evidence for statistical discrimination in favor of lesbians is mixed, however (Weichselbaumer 2003; Baert 2014).

Wage and Income Differences

Another traditional approach to testing for discrimination is to compare the earnings of LGBTQ to non-LGBTQ people, holding relevant observable factors constant, although wage differences could also reflect differences in unobserved characteristics rather than direct discrimination. In studying wage or earnings gaps for LGBTQ people, the convention has been to make sexual orientation comparisons within groups of men or women (typically using survey questions that did not allow identification of transgender respondents), while holding other influences on wages or incomes constant. To our knowledge, no existing study of population-based data has access to information about whether LGBTQ people have disclosed their identity to people in the workplace.

Reviews of the international body of research that includes data from several advanced economies has found a fairly consistent pattern for men: gay/bisexual men earned less than heterosexual men with the same education, age (or potential experience), race, marital status, geographic location, and other controls. For example, Klawitter (2015) found in a meta-analysis that on average gay/bisexual men earned 11 percent less than heterosexual men with the same characteristics. More recent reviews, notably Valfort (2017), continued to find negative earnings gaps for gay/bisexual men, as have most other studies published since then (Burn 2019; Aksoy, Carpenter, and Frank 2018). Some studies also found larger negative earnings gaps for bisexual men than for gay men (Aksoy, Carpenter, and Frank 2018).

The earnings patterns for women of different sexual orientations differ from those for men. Both Klawitter (2015) and Valfort (2017) found that earnings for lesbian/bisexual women were often higher than for heterosexual women (however, all groups of cisgender women, regardless of sexual orientation, tended to earn less than groups of cisgender men). Klawitter's meta-analysis found an average lesbian wage premium of 9 percent, and studies using data from the 1990 and 2000 Census showed a range of 7–8 percent higher earnings for women in same-sex couples. However, the measured differences for women varied widely across studies, ranging from −25 percent to 43 percent. In US studies, controlling for hours and weeks worked reduced the premium seen for lesbians. Recent studies of US data find mixed patterns: for example, Carpenter and Eppink (2017) showed higher earnings for lesbians and Martell (2019) showed lower earnings. A limitation of this literature—common to most studies of the gender gap in wages—is the difficulty in credibly accounting for the endogeneity of labor force participation, which is particularly relevant for comparisons of labor market outcomes between lesbian and heterosexual women.

Questions about gender identity rarely appear on representative surveys, making similar comparisons for transgender people difficult. Carpenter, Eppink,

and Gonzales (2020) found in the 35 states with information on gender identity in the Behavioral Risk Factor Surveillance System data that transgender-identified people had lower household incomes and lower employment rates than cisgender men with similar observed characteristics. Another study matched Dutch tax records and population registries to compare transgender people pre- and post-transition (Geijtenbeek and Plug 2018), finding that transgender female-to-male individuals earned much less per hour before transitioning than cisgender men, and their earnings changed very little after their transitions. In contrast, transgender, male-to-female individuals had similar hourly earnings compared to cisgender men when they were pre-transition, but saw a large drop in earnings and hours post-transition.

Discrimination or a Different Household Division of Labor?

The common findings of a negative wage gap for gay and bisexual men but a positive wage gap for lesbian and bisexual women raise obvious questions. Previous research in economics and sociology has found support for both discrimination and family decisions as explanations for wage and income gaps.

The audit studies mentioned earlier support the hypothesis of discrimination against LGBQ cisgender men and women and against transgender people. In addition, some studies of wage or employment gaps include variables that should capture some aspect of discrimination, such as statewide measures of nondiscrimination policies and public attitudes. The wage gap for men in same-sex couples is lower in states with a lower degree of prejudice (Burn 2019). States with nondiscrimination laws that include sexual orientation have somewhat lower earnings gaps for gay men (Burn 2018; Martell 2013). Tilcsik (2011) also found less differential treatment of gay male applicants when an employer was located in a state with a nondiscrimination law. When states opened up marriage to same-sex couples, men and women in same-sex couples worked more hours, and wages increased for men in same-sex couples, possibly reflecting a decline in discrimination (Sansone 2019a). Studies of self-reports of unfair treatment as well as discrimination complaints filed also offer evidence consistent with discrimination (Badgett, Baumle, and Boutcher 2020; Cech and Rothwell 2020).

Household structure of LGBTQ people might also contribute to wage gaps if people who partner with (or expect to partner with) a person of the same sex make different decisions about human capital investments and labor force attachment than those who plan to partner with a different-sex partner (Badgett 1995a; Antecol and Steinberger 2013; Black, Sanders, and Taylor 2007). Klawitter's (2015) meta-analysis found the largest wage differences in studies of data on same-sex couples, but the evidence on this point since then is mixed. In a UK study, only coupled people showed the gay-male penalty/lesbian-premium pattern (Aksoy, Carpenter, and Frank 2018), but a similar US study did not find that pattern (Carpenter and Eppink 2017).

Some studies have offered a Becker-style story of the household division of labor for people in same-sex couples that might differ by sex (Becker 1991). If gay men do not expect to support a partner and children, they might invest less in

unobserved human capital than heterosexual men do. But in most datasets, gay men have higher levels of education (as noted earlier), so it seems unlikely that unobserved human capital would be much lower for them. The household division of labor hypothesis could be a better explanation for the positive wage gap for lesbians. Although evidence from labor market and time-use studies suggests that female same-sex couples specialize to some extent, particularly when they have children, lesbian women might be more committed to the paid labor market because they are not likely to have a higher earning (male) partner to provide for them (Antecol and Steinberger 2013; Hansen, Martell, and Roncolato 2019). As a result, lesbians may invest more in human capital (education and labor force experience) that will raise their wages compared to heterosexual women. Because the data used in the wage gap studies typically include only measures of education but not actual experience, these studies might underestimate the advantage that lesbians have in actual experience that could explain their higher earnings.

Economists have used different strategies to capture lesbians' possible differences in labor market commitment and experience. Several studies have used an interaction term for being lesbian and potential experience and have found that the return on a year of potential experience (age minus years of education minus five), is higher for lesbians than for heterosexual women, as we would expect if potential experience is a better measure of actual experience for lesbians (Martell 2019). The lesbian wage premium becomes smaller or even negative when including that interaction. Also, the wage premium for women in same-sex couples is larger for those who were never married to men (Daneshvary, Waddoups, and Wimmer 2009) and those who are older (Martell 2019). Taken as a whole, this evidence suggests that lesbians have higher earnings in part because of a greater commitment to the paid labor force, an adaptation that could also offset both their earnings disadvantage as women and the potential negative effect of discrimination.

Changes over Time

We know little about whether rising acceptance of LGBTQ people has translated into changes in economic status. Klawitter's (2015) meta-analysis found that estimates of the wage gap for gay/bisexual men and the wage premium for lesbian/bisexual women are decreasing over time, but the trend is statistically insignificant after controlling for other characteristics of the studies. Other studies seeking to study the decline in the wage gap over time have been limited by small sample sizes and varying measures of sexual orientation.

To update earlier approaches with a larger dataset, we estimate wage gaps by year using the American Community Survey, which offers large samples of women and men in same-sex couples each year from 2000–2018. We restrict the respondents' age to be between 25 and 65 in this analysis, because we are focusing on labor market outcomes.[15] We focus on trends in wages, as trends in labor force

---

[15] Table C1 in the online Appendix reports detailed sample sizes by year, sex, and couple type. Independently of this study, Jepsen and Jepsen (2020) present related analyses.

participation were similar across couple types. To look at the trend in wage gaps, we estimate the inflation-adjusted hourly wage by dividing a respondent's annual wage and salary income by an estimate of hours worked in that year.

Panel A in Figure 3 presents wage gaps adjusted for demographic controls (age, race, ethnicity, education, citizenship, and disability) and state fixed-effects going back to 2000 for women in same-sex couples who usually worked full-time.[16] The lesbian premium fell from 10 percent in 2000, half the 20 percent gap in 1990 (Klawitter and Flatt 1998), to nearly zero by 2018.[17] A similar exercise for men in Panel B of Figure 3 shows a wage gap for men in same-sex couples that is negative every year, with no obvious trend, other than being much closer to zero than the estimated 26 percent gap in 1990 (Klawitter and Flatt 1998).

Our earlier discussion of data quality found that over time the sample composition of same-sex couples has become younger and more racially diverse, possibly because these groups are now more willing to report being in a same-sex couple. Further research is needed to assess the extent and impact of reporting bias and other factors on the diverging wage gaps for men and women.

Other Kinds of Discrimination

Several studies have found that lesbian, gay, and bisexual people work in different occupations than heterosexual people. The extent of occupational sorting for lesbian, gay and bisexual people is associated with more tolerant work settings and places where disclosure is less risky (Plug, Webbink, and Martin 2014). In addition, gender plays a smaller role in shaping the occupational positions of lesbian, gay and bisexual people than for heterosexual people (Del Río and Alonso-Villar 2019), which could be the result of gender stereotypes making it harder for lesbian, gay, and bisexual people to be accepted into occupations that are more traditional for people of their sex (Drydakis 2015; Tilcsik 2011). Other research found that LGB people were less likely to reach upper-level managerial jobs, suggesting the possibility that a glass ceiling holds back LGBTQ progress up job ladders (Aksoy et al. 2019).

Other workplace-related differentials are also apparent. In the United States, health insurance benefits offered by employers traditionally covered different-sex spouses of employees, but not necessarily same-sex partners. Although we are not aware of studies in the economics literature that have used large representative datasets to study transgender employees' workplace experiences, the lack of gender transition-related care in employer health benefits and issues such as access

---

[16] As discussed in Section C.1 of the online Appendix, the American Community Survey 2000–2007 have higher rates of misclassification errors. Nevertheless, we have decided to include here observations from these years to estimate longer trends in the wage differentials.

[17] As shown in Figures C5–C11, similar trends are observed when examining the raw wage gap, as well as when excluding state fixed effects, when including part-time workers, when including outliers in the wage distribution, when replicating the analysis without survey weights, when excluding individuals in different-sex unmarried couples, or when not adjusting wages for inflation.
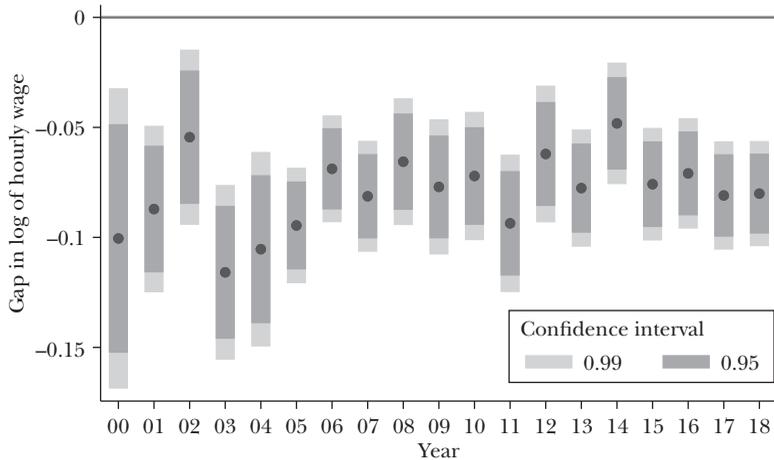
Figure 3

Gap in Hourly Wage for Individuals in Same-Sex versus Individuals in Different-Sex Couples

Panel A. Full-time women workers, with demographic controls and state fixed effects
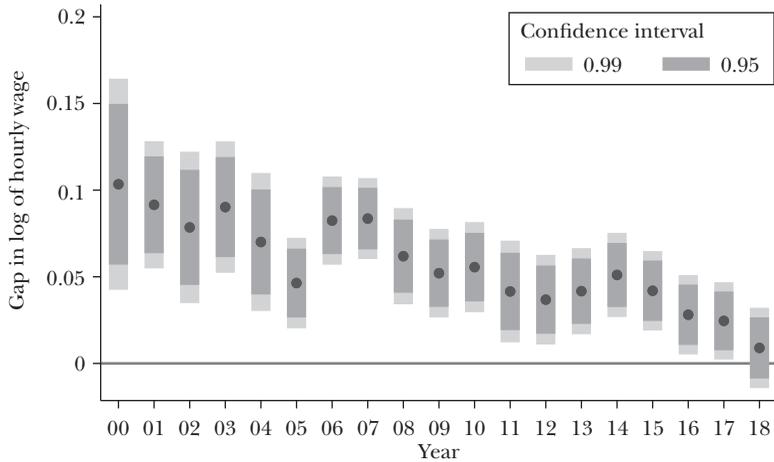
Panel B. Full-time men workers, with demographic controls and state fixed effects

Source: American Community Survey 2000–2012
Notes: These figures report the estimated gap between women (men in Panel B) in same-sex couples and women (men) in different-sex couples from 19 different regressions, one for each year. The dependent variable is the logarithm of the respondent's total pre-tax wage and salary income in the 12 months preceding the American Community Survey interview divided by the estimated number of hours worked in the same 12 months. All wages have been adjusted for inflation using the FRED Consumer Price Index for All Urban Consumers (All Items). Only respondents age 25–64 have been considered. Respondents whose hourly wage was above the 99th percentile of the hourly wage distribution for women (men) in same-sex and different-sex couples have been excluded. Only respondents with a positive hourly wage and working at least 40h/week have been included in the analysis. Both married and unmarried couples included in this sample. Weighted regressions using person weights. Confidence intervals computed using heteroskedasticity-robust standard errors.

to sex-segregated facilities and identification documents have been documented (National Academies of Sciences, Engineering, and Medicine 2020).

Housing discrimination is another growing research area. Several audit studies have tested the treatment of LGBTQ and non-LGBTQ applicants while searching for rental housing, finding evidence that prospective landlords discriminate against LGBTQ applicants (Ahmed and Hammarstedt 2009; Levy et al. 2017; Schwegman 2019). Other analyses of outcomes for actual mortgage loan applications finds evidence of discrimination against LGBTQ people in mortgage lending (Sun and Gao 2019).

Discussion and Conclusion

New and better data, taken together with the evolution of cross-jurisdiction law and rules affecting the LGBTQ population, have expanded the possibilities for high-quality studies in the area of LGBTQ economics. Continuing methodological work on measures of sexual orientation and gender identity will improve the value of these surveys for studying LGBTQ people (National Academies of Sciences Engineering and Medicine 2020). There are also a number of important research areas for economists.

First, more research is needed on transgender populations. We are not aware of any studies that use nationally representative samples to examine economic outcomes for transgender people, and we aware of only one study with nationally representative samples looking at economic patterns before and after transitions (Geijtenbeek and Plug 2018). This will likely require large surveys, such as the Current Population Survey or the American Community Survey, to include questions on gender identity, as Canada plans for its 2021 Census (LeBlanc 2020). Alternatively, as states increasingly allow individuals to change gender markers on legal documents, administrative data linkages (for example, with state unemployment insurance records) could provide evidence on outcomes for transgender people.

Second, previous research from sociology and psychology suggests the possibility of racial and ethnic heterogeneity in outcomes for sexual and gender minorities (Pedulla 2014; Schwegman 2019). However, small sample sizes have hindered the study of racial or ethnic differences in the earnings effects of sexual orientation. We carried out some cross-tabulations using our largest source of data—from the American Community Survey—to document socioeconomic and demographic outcomes by gender, couple type, and race (Table C13 in the online Appendix). While white and Hispanic women in same-sex couples have much higher education, labor force participation, and full-time employment rates than their same-race female counterparts in different-sex couples, these differences across couple type are much smaller or absent for Black and Asian women. Differences related to age, cohort, nationality, religion, geographic location, or rural/urban status within the LGBTQ population are also worthy of exploration.

Third, more research is needed concerning public policies relating to LGBTQ populations. While a literature on the effects of marriage equality is growing (as described above), more research is needed to understand the effects of legal access to same-sex marriage on savings and investment behavior, wealth, family formation (and divorce rates), and physical and mental health. Future research should explore the effect of the 2020 US Supreme Court ruling in *Bostock v. Clayton County* that employment discrimination based on sexual orientation and gender identity is illegal. Economists should also study the causes and consequences of policies that particularly affect LGBTQ individuals, such as health insurance reforms, religious freedom exemptions from nondiscrimination policies, or bills requiring students to use the bathroom consistent with their sex assigned at birth.

Fourth, there is a need for research on LGBTQ people in developing countries (Badgett 2020), where policy changes have gone in very different directions. On one side, India's Supreme Court decriminalized homosexuality in 2018, Taiwan legalized same-sex marriage in 2019, and Costa Rica introduced marriage equality in 2020. In contrast, anti-LGBTQ laws have been enacted in Hungary, Poland, Russia, Tanzania, and Uganda (Mendos 2019). More knowledge about the economic conditions and challenges faced by LGBTQ people, including the economic impact of these LGBTQ policy changes (Badgett, Waaldijk, and Rodgers 2019; Badgett 2020), could provide evidence to guide future decisions by development agencies and government actors.

Among the fields of economics, labor and demographic economists have been leaders in studying LGBTQ people. However, there are numerous research opportunities for health economists, public economists, development economists, economic historians, and macroeconomists to contribute their expertise to research questions addressing this important population.

References

Acquisti, Alessandro, and Christina Fong. 2020. "An Experiment in Hiring Discrimination via Online Social Networks." *Management Science* 66 (3): 1005–24.

Ahmed, Ali M., and Mats Hammarstedt. 2009. "Detecting Discrimination against Homosexuals: Evidence from a Field Experiment on the Internet." *Economica* 76 (303): 588–97.

Aksoy, Cevat G., Christopher S. Carpenter, and Jeff Frank. 2018. "Sexual Orientation and Earnings: New Evidence from the United Kingdom." *Industrial and Labor Relations Review* 71 (1): 242–72.

Aksoy, Cevat Giray, Christopher S. Carpenter, Jeff Frank, and Matt L. Huffman. 2019. "Gay Glass Ceilings : Sexual Orientation and Workplace Authority in the UK." *Journal of Economic Behavior & Organization* 159: 167–80.

Aksoy, Cevat G., Christopher S. Carpenter, Ralph De Haas, and Kevin D. Tran. 2020. "Do Laws Shape Attitudes? Evidence from Same-Sex Relationship Recognition Policies in Europe." *European Economic Review* 124.

Antecol, Heather, and Michael D. Steinberger. 2013. "Labor Supply Differences between Married Heterosexual Women and Partnered Lesbians: A Semi-Parametric Decomposition Approach." *Economic Inquiry* 51 (1): 783–805.

Arrow, Kenneth. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees, 3–33. Princeton: Princeton University Press.

Ash, Michael A., and M.V. Lee Badgett. 2006. "Separate and Unequal: The Effect of Unequal Access to Employment-Based Health Insurance on Same-Sex and Unmarried Different-Sex Couples." *Contemporary Economic Policy* 24 (4): 582–99.

Badgett, M.V. Lee. 1995a. "The Wage Effects of Sexual Orientation Discrimination." *ILR Review* 48 (4): 726–39.

Badgett, M.V. Lee. 1995b. "Gender Sexuality and Sexual Orientation: All in the Feminist Family?" *Feminist Economics* 1 (1): 121–39.

Badgett, M. V. Lee. 2009. "The Economic Value of Marriage for Same-Sex Couples." *Drake Law Review* 58: 1081–1120.

Badgett, M.V. Lee. 2018. "Left Out? Lesbian, Gay, and Bisexual Poverty in the U.S." *Population Research and Policy Review* 37 (5): 667–702.

Badgett, M.V. Lee. 2020. *The Economic Case for LGBT Equality: Why Fair and Equal Treatment Benefits Us All.* Boston, MA: Beacon Press.

Badgett, M.V. Lee, Amanda Baumle, and Steven Boutcher. 2020. "New Research on Sexual Orientation and Gender Identity Discrimination: Effect of State Policy on Charges Filed at the EEOC." *Journal of Homosexuality* 67 (8): 1135–44.

Badgett, M.V. Lee, Kees Waaldijk, and Yana van der Meulen Rodgers. 2019. "The Relationship between LGBT Inclusion and Economic Development: Macro-Level Evidence." *World Development* 120: 1–14.

Baert, Stijn. 2014. "Career Lesbians. Getting Hired for Not Having Kids?" *Industrial Relations Journal* 45 (6): 543–61.

Becker, Gary S. 1991. *A Treatise on the Family.* Cambridge, MA: Harvard University Press.

Black, Dan A., Gary Gates, Seth Sanders, and Lowell Taylor. 2000. "Demographics of the Gay and Lesbian Population in the United States: Evidence from Available Systematic Data Sources." *Demography* 37 (2): 139–54.

Black, Dan A., Gary Gates, Seth Sanders, and Lowell Taylor. 2002. "Why Do Gay Men Live in San Francisco?" *Journal of Urban Economics* 51 (1): 54–76.

Black, Dan A., Seth G. Sanders, and Lowell J. Taylor. 2007. "The Economics of Lesbian and Gay Families." *Journal of Economic Perspectives* 21 (2): 53–70.

Blewett, Lynn A., Julia A. Rivera Drew, Miriam L. King, and Kari C.W. Williams. 2019. "IPUMS Health Surveys: National Health Interview Survey." IPUMS Health Surveys 6.4. https://doi.org/10.18128/D070.V6.4. (accessed April 20, 2020).

Boertien, Diederik, and Fabrizio Bernardi. 2019. "Same-Sex Parents and Children's School Progress: An Association that Disappeared Over Time." *Demography* 56 (2): 477–501.

Burn, Ian. 2018. "Not All Laws Are Created Equal: Legal Differences in State Non-Discrimination Laws and the Impact of LGBT Employment Protections." *Journal of Labor Research* 39 (4): 462–97.

Burn, Ian. 2019. "The Relationship between Prejudice and Wage Penalties for Gay Men in the United States." *ILR Review* 73 (3).

Carpenter, Christopher S. 2020. "The Direct Effects of Legal Same-Sex Marriage in the United States: Evidence From Massachusetts." *Demography* 57 (5): 1787–1808.

Carpenter, Christopher S., and Samuel T. Eppink. 2017. "Does It Get Better? Recent Estimates of Sexual Orientation and Earnings in the United States." *Southern Economic Journal* 84 (2): 426–41.

Carpenter, Christopher S., Samuel T. Eppink, and Gilbert Gonzales. 2020. "Transgender Status, Gender Identity, and Socioeconomic Outcomes in the United States." *ILR Review* 73 (3): 573–99.

Carpenter, Christopher S., Samuel T. Eppink, Gilbert Gonzales, and Tara McKay. Forthcoming. "Effects

of Access to Legal Same-Sex Marriage on Marriage and Health: Evidence from BRFSS." *Journal of Policy Analysis and Management.*

Carpenter, Christopher, and Gary J. Gates. 2008. "Gay and Lesbian Partnership: Evidence from California." *Demography* 45 (3): 573–90.

CDC. 2019. Behavioral Risk Factor Surveillance System (BRFSS). https://www.cdc.gov/brfss/ (accessed September 5, 2019)

Cech, Erin A., and William R. Rothwell. 2020. "LGBT Workplace Inequality in the Federal Workforce: Intersectional Processes, Organizational Contexts, and Turnover Considerations." *ILR Review* 73 (1): 25–60.

Chen, Shuai, and Jan C Van Ours. 2018. "Subjective Well-Being and Partnership Dynamics: Are Same-Sex Relationships Different?" *Demography* 55 (6): 2299–2320.

Ciscato, Edoardo, Alfred Galichon, and Marion Goussé. 2020. "Like Attract Like? A Structural Comparison of Homogamy across Same-Sex and Different-Sex Households." *Journal of Political Economy* 128 (2): 740–81.

Coffman, Katherine B., Lucas C. Coffman, and Keith M. Marzilli Ericson. 2017. "The Size of the LGBT Population and the Magnitude of Antigay Sentiment Are Substantially Underestimated." *Management Science* 63 (10): 3168–86.

Dahlhamer, James M., Adena M. Galinsky, Sarah S. Joestl, and Brian W. Ward. 2014. "Sexual Orientation in the 2013 National Health Interview Survey: A Quality Assessment." *Vital and Health Statistics* 2 (169): 1–24.

Daneshvary, Nasser, C. Jeffrey Waddoups, and Bradley S. Wimmer. 2009. "Previous Marriage and the Lesbian Wage Premium." *Industrial Relations* 48 (3): 432–53.

Dee, Thomas S. 2008. "Forsaking All Others? The Effects of Same-Sex Partnership Laws on Risky Sex." *The Economic Journal* 118 (530): 1055–78.

Dillender, Marcus. 2014. "The Death of Marriage? The Effects of New Forms of Legal Recognition on Marriage Rates in the United States." *Demography* 51 (2): 563–85.

Drydakis, Nick. 2015. "Sexual Orientation Discrimination in the United Kingdom's Labour Market: A Field Experiment." *Human Relations* 68 (11): 1769–96.

England, Paula, Emma Mishel, and Mónica L. Caudillo. 2016. "Increases in Sex with Same-Sex Partners and Bisexual Identity across Cohorts of Women (but Not Men)." *Sociological Science* 3: 951–70.

Federal Interagency Working Group. 2016. *Evaluations of Sexual Orientation and Gender Identity Survey Measures: What Have We Learned?* Federal Interagency Working Group.

Flores, Andrew R., and Scott Barclay. 2016. "Backlash, Consensus, Legitimacy, or Polarization: The Effect of Same-Sex Marriage Policy on Mass Attitudes." *Political Research Quarterly* 69 (1): 43–56.

Gates, Gary J. 2011. "How Many People Are Lesbian, Gay, Bisexual, and Transgender?" *The Willians Institute.*

Gates, Gary J. 2015a. "Marriage and Family: LGBT Individuals and Same-Sex Couples." *The Future of Children* 25 (2): 67–87.

Gates, Gary J. 2015b. *Demographics of Married and Unmarried Same-Sex Couples: Analyses of the 2013 American Community Survey.* Los Angeles: The Williams Institute.

Geijtenbeek, Lydia, and Erik Plug. 2018. "Is There a Penalty for Registered Women? Is There a Premium for Registered Men? Evidence from a Sample of Transsexual Workers." *European Economic Review* 109: 334–47.

Gonzales, Gilbert, and Lynn A. Blewett. 2014. "National and State-Specific Health Insurance Disparities for Adults in Same-Sex Relationships." *American Journal of Public Health* 104 (2): 95–105.

Granberg, Mark, Per A. Andersson, and Ali Ahmed. 2020. "Hiring Discrimination against Transgender People: Evidence from a Field Experiment." *Labour Economics* 65.

Hamermesh, Daniel S, and Scott M. Delhommer. Forthcoming. "Same-Sex Couples and the Gains to Marriage: The Importance of the Legal Environment." *Journal of Policy Analysis and Management.*

Hansen, Mary Eschelbach, Michael E. Martell, and Leanne Roncolato. 2019. "A Labor of Love: The Impact of Same-Sex Marriage on Labor Supply." *Review of Economics of the Household* 18 (2): 265–83.

Herek, Gregory M., Aaron T. Norton, Thomas J. Allen, and Charles L. Sims. 2010. "Demographic, Psychological, and Social Characteristics of Self-Identified Lesbian, Gay, and Bisexual Adults in a US Probability Sample." *Sexuality Research and Social Policy* 7 (3): 176–200.

Herman, Jody L., Andrew R. Flores, Taylor N.T. Brown, Bianca D.M. Wilson, and Kerith J. Conron. 2017. *Age of Individuals Who Identify as Transgender in the United States.* Los Angeles: The Williams Institute.

Institute of Medicine. 2011. *The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a*

Foundation for Better Understanding. Washington, D.C.: The National Academies Press.

Jepsen, Christopher A., and Lisa K. Jepsen. 2002. "An Empirical Analysis of the Matching Patterns of Same-Sex and Opposite-Sex Couples." *Demography* 39 (3): 435–53.

Jepsen, Christopher A., and Lisa K. Jepsen. 2020. Convergence Over Time or Not? U.S. Wages by Sexual Orientation, 2001–2018." *IZA Discussion Paper* 13495.

Kinsey, Alfred C., Wardell B. Pomeroy, and Clyde E. Martin. 1948. *Sexual Behavior in the Human Male*. Bloomington, IN: Indiana University Press.

Kinsey, Alfred C., Wardell B. Pomeroy, Clyde E. Martin, and Paul H. Gebhard. 1953. *Sexual Behavior in the Human Female*. Bloomington, IN: Indiana University Press.

Klawitter, Marieka. 2015. "Meta-Analysis of the Effects of Sexual Orientation on Earnings." *Industrial Relations: A Journal of Economy and Society* 54 (1): 4–32.

Klawitter, Marieka M., and Victor Flatt. 1998. "The Effects of State and Local Antidiscrimination Policies on Earnings for Gays and Lesbians." *Journal of Policy Analysis and Management* 17 (4): 658–86.

Laumann, Edward O., John H. Gagnon, Robert T. Michael, and Stuart Michaels. 2000. *The Social Organization of Sexuality: Sexual Practices in the United States*. Chicago: University of Chicago Press.

LeBlanc, Daniel. 2020. "Ottawa Adding New Census Questions on Gender, Indigenous People, Linguistic and Ethnic Minorities." *The Globe and Mail*, July 16, https://www.theglobeandmail.com/politics/article-ottawa-adding-new-census-questions-on-gender-indigenous-canadians/.

Levy, Diane K., Douglas Wissoker, Claudia L. Aranda, Brent Howell, Rob Pitingolo, Sarale Sewell, and Rob Santos. 2017. *A Paired-Testing Pilot Study of Housing Discrimination against Same-Sex Couples and Transgender Individuals*. Washington, DC: Urban Institute.

Mackenzie-Liu, Mattie, David J. Schwegman, and Leonard M. Lopoo. Forthcoming. "Do Foster Care Agencies Discriminate Against Gay Couples? Evidence from a Correspondence Study." *Journal of Policy Analysis and Management*.

MAP. 2019. "Where We Call Home: LGBT People in Rural America." Boulder, CO: Movement Advancement Project.

Martell, Michael E. 2013. "Do ENDAs End Discrimination for Behaviorally Gay Men?" *Journal of Labor Research* 34 (2): 147–69.

Martell, Michael E. 2019. "Age and the New Lesbian Earnings Penalty." *International Journal of Manpower* 41 (6): 649–70.

Mendos, Lucas Ramón. 2019. *State-Sponsored Homophobia 2019: Global Legislation Overview Update*. Geneva: ILGA.

Meyer, Ilan H., Taylor N.T. Brown, Jody L. Herman, Sari L. Reisner, and Walter O. Bockting. 2017. "Demographic Characteristics and Health Status of Transgender Adults in Select US Regions: Behavioral Risk Factor Surveillance System, 2014." *American Journal of Public Health* 107 (4): 582–89.

Miller, Joshua J., and Kevin A. Park. 2018. "Same-Sex Marriage Laws and Demand for Mortgage Credit." *Review of Economics of the Household* 16 (2): 229–54.

Miller, Kristen, and J. Michael Ryan. 2011. *Design, Development and Testing of the NHIS Sexual Identity Question*. Hyattsville, MD: National Center for Health Statistics.

National Academies of Sciences, Engineering, and Medicine. 2020. *Understanding the Well-Being of LGBTQI+ Populations*. Washington, D.C.: National Academies of Sciences, Engineering, and Medicine.

Neumark, David. 2018. "Experimental Research on Labor Market Discrimination." *Journal of Economic Literature* 56 (3): 799–866.

Newport, Frank. 2018. "In U.S., Estimate of LGBT Population Rises to 4.5%." *Gallup,* May 22 https://news.gallup.com/poll/234863/estimate-lgbt-population-rises.aspx.

OECD. 2019. *Society at a Glance 2019: OECD Social Indicators*. Paris: OECD.

OECD. 2020. *Over the Rainbow? The Road to LGBTI Inclusion*. Paris: OECD.

Ofosu, Eugene K., Michelle K. Chambers, Jacqueline M. Chen, and Eric Hehman. 2019. "Same-Sex Marriage Legalization Associated with Reduced Implicit and Explicit Antigay Bias." *Proceedings of the National Academy of Sciences of the United States of America* 116 (18): 8846–51.

Oreffice, Sonia. 2011. "Sexual Orientation and Household Decision Making.: Same-Sex Couples' Balance of Power and Labor Supply Choices." *Labour Economics* 18 (2): 145–58.

Patterson, Charlotte J. 2005. "Lesbians and Gay Parents and Their Children: Summary of Research Findings." In *Lesbian & Gay Parenting*, 5–22. Washington, DC: American Psychological Association.

Pearson, Jennifer, and Lindsey Wilkinson. 2017. "Same-Sex Sexuality and Educational Attainment: The Pathway to College." *Journal of Homosexuality* 64 (4): 538–76.

Pedulla, David S. 2014. "The Positive Consequences of Negative Stereotypes: Race, Sexual Orientation, and the Job Application Process." *Social Psychology Quarterly* 77 (1): 75–94.

Plug, Erik, Dinand Webbink, and Nick Martin. 2014. "Sexual Orientation, Prejudice, and Segregation." *Journal of Labor Economics* 32 (1): 123–59.

Pornhub Insights. 2017. "Pornhub Pride." *Pornhub*, June 22. https://www.pornhub.com/insights/gay-porn-pride.

Ridolfo, Heather, Kristen Miller, and Aaron Maitland. 2012. "Measuring Sexual Identity Using Survey Questionnaires: How Valid Are Our Measures?" *Sexuality Research and Social Policy* 9: 113–24.

Río, Coral Del, and Olga Alonso-Villar. 2019. "Occupational Achievements of Same-Sex Couples in the United States by Gender and Race." *Industrial Relations* 58 (4): 704–31.

Robertson, Ronald E., Felix W. Tran, Lauren N. Lewark, and Robert Epstein. 2018. "Estimates of Non-Heterosexual Prevalence: The Roles of Anonymity and Privacy in Survey Methodology." *Archives of Sexual Behavior* 47 (4): 1069–84.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2020. "IPUMS USA 10.0." IPUMS, Minneapolis, MN. https://doi.org/10.18128/D010.V10.0. (accessed April 21, 2020).

Sansone, Dario. 2019a. "Pink Work: Same-Sex Marriage, Employment and Discrimination." *Journal of Public Economics* 180.

Sansone, Dario. 2019b. "LGBT Students: New Evidence on Demographics and Educational Outcomes." *Economics of Education Review* 73.

Sansone, Dario, and Christopher S. Carpenter. 2020. "Turing's Children: Representation of Sexual Minorities in STEM." *PLOS ONE* 15 (11).

Schwegman, David. 2019. "Rental Market Discrimination against Same-Sex Couples: Evidence from a Pairwise-Matched Email Correspondence Test." *Housing Policy Debate* 29 (2): 250–72.

Stephens-Davidowitz, Seth. 2017. *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us about Who We Really Are.* New York: Harper Collins.

Sun, Hua, and Lei Gao. 2019. "Lending Practices to Same-Sex Borrowers." *Proceedings of the National Academy of Sciences of the United States of America* 116 (19): 9293–9302.

Taylor, Jami K., Donald P. Haider-Markel, and Benjamin Rogers. 2019. "Toward a New Measure of State-Level LGBT Interest Group Strength." *State Politics & Policy Quarterly* 19 (3): 334–50.

Tilcsik, András. 2011. "Pride and Prejudice: Employment Discrimination against Openly Gay Men in the United States." *American Journal of Sociology* 117 (2): 586–626.

Trandafir, Mircea. 2015. "Legal Recognition of Same-Sex Couples and Family Formation." *Demography* 52: 113–51.

Valfort, Marie-Anne. 2017. "LGBTI in OECD Countries: A Review." OECD Social, Employment and Migration Working Paper 198.

Weichselbaumer, Doris. 2003. "Sexual Orientation Discrimination in Hiring." *Labour Economics* 10 (6): 629–42.

The Ways of Corruption in Infrastructure: Lessons from the Odebrecht Case

Nicolás Campos, Eduardo Engel, Ronald D. Fischer, and Alexander Galetovic

I n 2010, the Swiss business school IMD chose Odebrecht, a Brazilian conglomerate, as the world's best family business. Odebrecht was chosen for the excellent performance of its companies, its continuous growth, and its social and environmental responsibility. Sales had quintupled between 2005 and 2009, and Odebrecht had become Latin America's largest engineering and construction company and ranked 18th worldwide among international contractors (*Engineering News-Record Magazine* 2009).

By 2015, however, Odebrecht chief executive Marcelo Odebrecht had been arrested on corruption charges. Nine months later he was sentenced to more than 19 years in prison. The Odebrecht case, as it came to be known, involved bribe payments in ten countries in Latin America and two countries in Africa. Deltan Dallagnol, lead prosecutor in Brazil, commented (as reported by Pressly 2018): "The Odebrecht case leaves you speechless. This case implicated almost one-third of Brazil's senators and almost half of all Brazil's governors. A single company paid bribes to 415 politicians and 26 political parties in Brazil. It makes the Watergate

■ *Nicolás Campos, Researcher, Espacio Público, Santiago, Chile. Eduardo Engel is Professor of Economics, University of Chile, and Director, Espacio Público, both in Santiago, Chile. Ronald D. Fischer is Professor of Industrial Engineering, University of Chile, and a Research Associate, ISCI (Instituto Sistemas Complejos de Ingeniería), both in Santiago, Chile. Alexander Galetovic is Senior Fellow, Universidad Adolfo Ibáñez, Santiago, Chile; Research Fellow, Hoover Institution, Stanford, California; and Research Associate, CRIEP (Centro di Ricerca Interuniversitario sull'Economia Pubblica), Padua. Italy. Their email addresses are ncamposbijit@gmail.com, rfischer@dii.uchile.cl, emraengel@gmail.com, and agaleto@ stanford.edu .*

scandal look like a bunch of kids playing in a sandbox." The US Department of Justice (2016) described the case as "the largest foreign bribery case in history."

The US Department of Justice prosecuted Odebrecht under the US Foreign Corrupt Practices Act of 1977, which prohibits paying bribes. The US Department of Justice had jurisdiction in the Odebrecht case, as it did in a number of other international bribery cases, because the company made payments from bank accounts in New York, and some meetings to negotiate bribes were held in Miami (as reported in Shield and Chavkin 2019). As shown in Table 1, Odebrecht is by far the largest case prosecuted in the 40-year history of the Foreign Corrupt Practices Act, both in terms of the profits obtained from corruption and the size of the fine. Indeed, Table 1 shows that Odebrecht's profits from corruption were as large as the combined profits of the remaining nine firms.

The Odebrecht Bribes

The Odebrecht case emerged as an offshoot of the Lava Jato ("Car Wash") investigation in Brazil. The Lava Jato case began as a minor investigation of money laundering by *doleiros*, black market foreign exchange dealers operating through car washes and gas stations. As investigations continued, links were found to Petrobras, the large Brazilian state-controlled oil company. In plea agreements, Petrobras executives confessed that between 2004 and 2012 they colluded with contractors to run a bid-rigging scheme that exchanged contracts for bribes. Contractors would pay bribes of 1–3 percent of the value of the contract, which were split between Petrobras executives, politicians, and political parties. As the Lava Jato investigation unfolded, it uncovered a separate corruption scheme run by the construction firm Odebrecht (and its petrochemical affiliate Braskem). As it turned out, Odebrecht had bribed about 600 politicians and public servants in ten Latin American countries to win the public bidding process of large infrastructure projects, and to renegotiate the projects at higher prices after winning them.

By 2006, bribery at Odebrecht had become so institutionalized that the company created the Division of Structured Operations (DSO), a stand-alone department dedicated to corruption. According to the plea agreement between the Odebrecht chief executive officer Marcelo Odebrecht and the US Department of Justice, the DSO specialized in buying influence through legal and illegal contributions to political campaigns and also in paying bribes to public officials and politicians. Within the DSO, three full-time executives and four experienced assistants were responsible for paying bribes to foreign accounts. Bribe payments followed a clear organizational flow. A contract manager would deal with potential bribe recipients—public officials and politicians—and reported to the country manager. The country manager could approve small bribes paid with local funds. Larger bribes were vetted by an executive reporting directly to the Odebrecht chief executive officer who often made the final decision.

Table 1

Top Ten Foreign Corrupt Practices Act Cases: Gross Profits from Bribes

(in millions of US dollars)

Case	Countries	Gross profits from bribes[a]	Amount of bribes paid	Total fine[c]	Countries to which fines were paid
Odebrecht (2001–2016)	Angola, Argentina, Brazil, Colombia, Dominican Republic, Ecuador, Guatemala, Mexico, Mozambique, Panama, Peru, Venezuela	3,336	788	2,600	Brazil, Switzerland, United States
Siemens (1996–2007)	Argentina, Bangladesh, China, Iraq, Israel, Mexico, Nigeria, Russia, Venezuela, Vietnam	1,100[b]	1,400[b]	1,600	Germany, United States
Societe Generale and Legg Mason (2004–2011)	Libya	523	91	860	France, United States
Keppel (2001–2014)	Brazil, Iraq	500	55	422	Brazil, Singapore, United States
Ericsson (2000–2017)	China, Djibouti, Indonesia, Kuwait, Saudi Arabia, Vietnam	458	150	1,060	United States
Telia (2007–2012)	Uzbekistan	457	331	965	Netherlands, Sweden, United States
Alstom (2000–2010)	Bahamas, Egypt, Indonesia, Saudi Arabia, Taiwan	296	75	860	United States
Teva (n.a.)	Mexico, Russia, Ukraine	221	n.a.	541	United States, Israel
Total (1995–2005)	Iran	150	60	398	United States
Fresenius (2009–2016)	Angola, Benin, Bosnia and Herzegovina, Burkina Faso, Cameroon, China, Gabon, Ivory Coast, Mexico, Morocco, Niger, Saudi Arabia, Senegal, Serbia and Montenegro, Spain, Turkey	140	30	232	United States

Note: [a]Gross profits from bribes and bribes paid were extracted from Stanford's Law School Foreign Corrupt Practices Act Clearinghouse Database at https://law.stanford.edu/foreign-corrupt-practices-act-clearinghouse-fcpac. Gross profits are profits before paying bribes. This table considers information up to January 2020. [b]Siemens paid more in bribes than the gross profits it made. This is consistent with the information contained in the complaint of US Securities and Exchange Commission 2008, p. 2. [c]Total fines were extracted from Stanford Law School Foreign Corrupt Practices Act Clearinghouse Database and the OECD (2019, p. 119).

Once a bribe was authorized, the Division of Structured Operations registered, managed, and made the payment through a network of shell companies, off-book transactions, and off-shore bank accounts. This included the Antigua subsidiary of Austria's Meinl Bank, bought for this purpose by Odebrecht. The DSO also used an independently funded parallel cash trove (called Caixa 2). As the US Department of Justice (2016) described the arrangements: "[T]o conceal its activities, the Division of Structured Operations utilized an entirely separate and off-book communications system . . . to communicate with one another and with outside financial operators . . . via secure emails and instant messages, using codenames and passwords." The DSO also used a bespoke information management system for bookkeeping and to track information flows.

Before the creation of the Division of Structured Operations, bribes in the construction sector in Latin America were usually paid in cash, which is inefficient, because some of the money in the suitcase "leaks." According to the executive who headed the DSO: "When you are working with cash that is off the books, it can disappear. So they needed someone who could guarantee it wouldn't disappear" (as reported by Smith, Valle, and Schmidt 2017). Moreover, the bribee must conceal and launder the cash. The DSO was designed to solve both "problems," with full-time employees making payments through a web of offshore entities via tax havens with strong banking secrecy laws. In such ways, the DSO increased the effectiveness of bribe payments and reduced their cost.

On December 21, 2016, 77 current and former Odebrecht executives signed plea agreements with the US Department of Justice and with Swiss and Brazilian authorities in exchange for leniency. Marcelo Odebrecht, the former CEO of Odebrecht, and two other executives were sentenced in Brazil to 19 years in prison for corruption, money laundering, and criminal association. However, the sentence was later reduced, and Marcelo Odebrecht was instead placed under house arrest.

As Table 2 shows, the plea agreement between Odebrecht and the US Department of Justice comprised bribes paid in ten Latin American countries (Argentina, Brazil, Colombia, Dominican Republic, Ecuador, Guatemala, Mexico, Panama, Peru, and Venezuela), and two African countries (Angola and Mozambique). For the first eight countries listed in Table 2, we were able to determine the initial cost estimate and the final cost after renegotiation of each project that Odebrecht undertook during the period covered by the agreement with the US Department of Justice. This includes 88 projects: 62 where Odebrecht paid bribes and 26 projects without bribes. The 88 projects were procured either as public works (68 projects) or as public-private partnerships (20 projects). Odebrecht also built 140 projects in Brazil, of which we were able to gather data on 105. Bribes were paid in 72 of them.

With the exception of Venezuela and Mozambique, the US Department of Justice was able to estimate gross profits made by paying bribes in each country—"any profit earned on a particular project for which a profit was generated as the result of a bribe payment" (the third column in Table 2). From Table 2,

Table 2

The Odebrecht Case: Basic Statistics

(in millions of US dollars)

Country	Bribes[a]	Gross profits from bribes[a]	Projects[b]	Projects with bribes[c]	Initial cost[d]	Cost after renegotiation[d]
Argentina (2007–2014)	35	278	6	5	4,141	13,343
Colombia (2009–2014)	11	50	4	3	1,828	2,134
Dominican Republic (2001–2014)	92	163	16	15	4,588	5,853
Ecuador (2007–2016)	33.5	116	10	7	3,466	4,074
Guatemala (2013–2015)	18	34	1	1	384	384
Mexico (2010–2014)	10.5	39	6	3	2,155	3,059
Panama (2010–2014)	59	175	20	13	8,839	10,391
Peru (2005–2014)	29	143	25	15	14,904	17,253
Brazil (2004–2016)	349	1,900	105	72	66,080	77,559
Total (2001–2016)	637	2,898	193	134	106,384	134,051
Angola (2006–2013)	50	261.7	n.a.	n.a.	n.a.	n.a.
Mozambique (2011–2014)	0.9	n.a.	n.a.	n.a.	n.a.	n.a.
Venezuela (2006–2015)	98	n.a.	n.a.	n.a.	n.a.	n.a.
Total (all countries) (2001–2016)	786	3,160	n.a.	n.a.	n.a.	n.a.

Note: [a]Bribes and gross profits from bribing were taken from the plea agreement between Odebrecht and the US Department of Justice. Gross profits from bribing means any incremental profit obtained because Odebrecht paid a bribe, gross of bribes. In projects where bribes exceeded the estimated profits, the Department of Justice set gross profits equal to the bribe. [b]The number of projects in each country was obtained from Odebrecht's annual reports. [c]The number of projects where Odebrecht paid a bribe was obtained from legal documents and press reports. [d]We estimated initial cost and cost after renegotiations for each project undertaken as follows. First, from Odebrecht's annual reports and the websites of Odebrecht's subsidiaries, we obtained the list of public infrastructure projects awarded each year to Odebrecht in each country. Second, for each country we defined a list of websites where we searched for documents mentioning the projects. Third, we divided the websites in two categories: official and media sources. Official sources contain documents issued by the government, judicial bodies, reports from the Comptroller General, reports from investigative commissions related to the Odebrecht case in the country, and information provided in annual company reports. Media sources included reports from investigative media and information provided by nongovernment organizations. Fourth, we downloaded all documents from official and media sources that mentioned a project. These documents included contracts, documents related to the bidding process, supplementary contracts, depositions of Odebrecht executives, legal documents from judicial bodies, media reports, and information provided by nongovernment organizations. Fifth, we reviewed each document searching for the initial and final cost. If two (or more) official sources provided different information for a project, preference was given to information from contracts (original and supplementary), and secondly to documents related to the tendering process. If two (or more) media sources provided different information for the same project, priority was given to information from investigative media sources. The increase in costs due to renegotiations was estimated in real terms.

Odebrecht paid $786 million in bribes and obtained gross profits equal to $3,160 million.[1] That is, Odebrecht made $3 in net profits for every $1 it paid in bribes.

The last two columns of Table 2 report the initial cost estimates and the cost increase after renegotiations. For the eight countries with complete data at the top panel of the table, cost increased by 40.1 percent after renegotiations, with substantial variation across countries and projects. For the 105 projects from Brazil for which we have data, costs increased by 17.4 percent.

The Odebrecht case had major economic and political consequences throughout Latin America. In many cases, large projects were suspended or abandoned due to anticorruption clauses in the contracts (de Michele, Prats, and Losada 2018). For example, construction of Gasoducto del Sur, a large pipeline duct in Peru that would transport natural gas from the Camisea fields to the south of the country, was suspended even though the generating plants that would use the gas had already been built. The IMF (2018, p. 21, box 2) estimates that the macroeconomic cost brought about by the Odebrecht case in Peru was of the order of 0.8 percent of GDP in 2017. Though there are no definitive estimates, several reports speculate that the Lava Jato and the associated Odebrecht case had a significant macroeconomic impact in Brazil as well: for example, the Lava Jato case has generated a suspension of projects worth approximately $27 billion (as reported by Pereira 2017).

The plea agreement between the US Department of Justice and Odebrecht triggered judicial investigations in several countries, leading to plea bargains and additional disclosures of political corruption. In Peru, President Pedro Pablo Kuczynski was forced to resign, and of the three previous Peruvian Presidents, Alan García committed suicide, Alejandro Toledo fled the country, and Ollanta Humala spent time in jail. In Brazil, former president Luis Inácio Lula da Silva spent 19 months in prison in connection with alleged bribe payments made by rival construction firm OAS; in Ecuador, former vice-president Jorge Glas was sentenced to six years in jail. The Odebrecht case may have weakened the confidence of the public in democracy and helped lead to the current wave of populism in Latin America. As Simon (2019) argued: "From Mexico to Brazil, the Odebrecht scandal helped push corruption to the center of public debate. It also bolstered a widespread revolt against political and business elites—a decisive element in most of the elections held in Latin America over the past two years."

---

[1] In the plea agreement, the US Department of Justice states that Odebrecht's gross profits from bribing were equal to $3,336 million and that bribes paid were $788 million. We cannot explain this discrepancy in the US Department of Justice numbers. In our discussion, we will use the country-level data from the table.

What Did Odebrecht Obtain in Exchange for Bribes?

Our examination of judicial documents and media reports provides details on the *quid pro quos* between Odebrecht and corrupt officials. They show that Odebrecht bribed to tailor auctions in its favor and to obtain favorable terms when renegotiating the contract after the projects were awarded.

Manipulation of Subjective Bid Criteria

The literature has argued extensively that subjective criteria in government bidding are prone to corruption, even when the tendering process is open (for example, see Huang and Xia 2019; Tran 2009; Burguet and Perry 2007; Burguet and Che 2004).

Odebrecht distorted the firm selection process in various ways. The evaluation of the technical expertise of participants was often biased in projects that were tendered competitively. For example, if the technical score was a weighted average of objective and subjective components, the weights would be chosen to favor Odebrecht. Alternatively, Odebrecht could be arbitrarily awarded the highest possible technical score, while its competitors received a lower score. In other cases, potential bidders were disqualified by setting technical requirements that only Odebrecht could meet.

As one example, consider the tender for the construction of the Trasvase Daule-Vinces reservoir in Ecuador. The final score was the weighted average of the technical score (55 percent) and the cost bid (45 percent). According to prosecutors, Odebrecht paid $6 million to Carlos Villamarín, the president of the tender commission, to ensure that the only rival of Odebrecht received a lower technical score and thus Odebrecht won the contract.[2]

A second example is the Poliducto Pascuales-Cuenca pipeline in Ecuador. Odebrecht paid $5 million to José Rubén Terán, a bribe intermediary, who distributed the bribes to Petroecuador's chief executive officer and to three other executives responsible for the auction. These executives tailored the technical requirements and the documents needed to prove experience so as to disqualify the other three bidders, and Odebrecht won the project.[3]

Finally, in the tender for the Santos Dumont Airport in Rio de Janeiro, Brazil, Odebrecht paid $3.8 million to the president and three board members of Infraero, the government agency in charge of airports. In exchange, these officers added financial requirements to the tender documents that disqualified six of the twelve companies that entered the tender. The remaining six companies were members

---

[2] See "Judicial Report of *Process N° 17721-2017-00222*," Procuraduría General del Estado, p. 6, Jan 29, 2019. Available at https://es.scribd.com/document/394930485/ReporteProceso-17721-2017-00222-Juicio-Asoc-Ilicita.

[3] See Deposition of José Conciencao Santos Filho, an Odebrecht executive, to Ecuadorian prosecutors, pp. 6–9, Available at https://www.dropbox.com/s/xho8srvafa3t6gf/Delaci%C3%B3n%20Jos%C3%A9%20C.%20Santos.pdf?dl=0.

of a cartel led by Odebrecht. After sham competition between the members of the cartel, Odebrecht won the tender.

Better Terms in Renegotiations

Bribing to obtain better terms in a renegotiation has also been mentioned as a *quid pro quo* in the research literature (for example, Guasch and Straub 2009). Nevertheless, direct evidence is hard to find, and most studies only search for a correlation between the frequency and size of renegotiations and an aggregate corruption indicator.

There exists ample evidence of renegotiations of infrastructure contracts that do not necessarily involve bribe payments. For example, Bajari, Houghton, and Tadelis (2014) examined 819 highway procurement contracts in California and found that the final price was, on average, 5.8 percent higher. This raises the question of whether the cost increases were larger in projects where Odebrecht paid bribes.

The first column of Table 3 shows the cost increase after renegotiations for projects with and without bribes. Data on the amount renegotiated comes from government agencies. We determined whether bribes were paid in each project doing a thorough search and review of legal records and media sources. As the first column of the table shows, in the 26 projects with no bribes, costs increased by a weighted average of 5.6 percent after renegotiations (simple average 16.3 percent). In contrast, in the 62 projects where Odebrecht paid bribes, costs rose by 70.8 percent after renegotiations. Thus, cost increases in renegotiations are about 12 times larger when Odebrecht paid a bribe. As a robustness check, the second column of the table repeats the computations using only legal documents. Under this stricter criterion, Odebrecht paid bribes in 45 projects. Now the amount renegotiated increases from 10.9 percent when there are no bribes, compared to 84.9 percent with bribes.

Data from the Odebrecht projects in Brazil for which we could obtain information also show that renegotiations were larger when bribes were paid, even though cost increases were smaller overall. Specifically, using legal and media sources to detect bribe payments, we find that renegotiations in projects with bribes led to a cost increase of 18.9 percent compared with 4.1 percent for projects without bribes (both weighted averages). If instead we consider simple averages, the percentages are 24.5 and 6.2 percent, respectively. Again, as a robustness check, if we only consider legal sources to determine whether bribes were paid, the above percentages are 18.8 versus 6.2 percent for weighted averages (or 24.6 versus 6.9 percent for simple averages).

To the best of our knowledge, the evidence presented above is the first to establish a direct link between bribe payments and the magnitude of contract renegotiations.

Our examination of judicial documents and media reports confirms that Odebrecht paid bribes in the expectation that it would renegotiate the contract to its advantage. Consider the Vía Costa Verde–Tramo Callao project in Peru. Under

Table 3
Cost Increase after Renegotiations

		Evidence of bribes	
		Legal or media	*Legal*
No bribes	Number of projects	26	43
	Simple average:	16.3%	23.3%
	Weighted average:	5.6%	10.9%
Bribes	Number of projects	62	45
	Simple average:	59.1%	68.6%
	Weighted average:	70.8%	84.9%

Source: Authors' calculations using data from the US Department of Justice, media, and investments as reported by government agencies.

Peruvian law, the Ministry of Finance sets a "reference cost" for any project and requires bids to be within 10 percent of the reference value—or else be rejected outright. Odebrecht, which had previously paid $4 million to Felix Moreno, the regional governor of Callao, asked him to increase the reference value. According to the plea agreement of an Odebrecht executive, Moreno pointed out that the Ministry of Finance would not acquiesce to a change in the reference value but promised to increase the project value in a subsequent contract renegotiation (Poder Judicial del Perú 2019). Eventually the contract was renegotiated eight times and the total cost increased by 55 percent from $106 million to $161 million.

As an example of renegotiations that added major works to the initial project, consider the Linea Noroeste aqueduct in the Dominican Republic. According to the Prosecutor of the Dominican Republic, Odebrecht acted through Ángel Rondón, a well-connected businessman, to bribe two successive Directors of the Water Works, and then also bribed Porfirio Bautista, President of the National Senate, which had to approve the budget for the additional works. Odebrecht paid $1.6 million to enlarge the project, increasing the value of the contract by $89 million (Poder Judicial de la República Dominicana 2018). The contract was renegotiated four times and its cost increased from $161 to $250 million.

An even more extreme example is the hydroelectric plant Pinalito in the Dominican Republic. Odebrecht bribed the Vice President of the Dominican Corporation of State-Owned Electric Companies (CDEEE), to add a fully independent project to the original contract (Poder Judicial de la República Dominicana 2018). The addition was the El Abanico-Constanza road, which increased the value of the contract by $88 million. The Pinalito contract was eventually renegotiated six times and the total cost increased from $131 million to $231 million.

At times, bribes were paid to circumvent the very controls meant to prevent opportunistic renegotiation. For example, consider the agreement between Odebrecht and the Ministry of Transport and Communications during the execution

Table 4
Reasons for Bribing

Tailored bidding process	Favorable renegotiations	Extortion	Number of projects
Yes	Yes	Yes	1
Yes	Yes	No	19
Yes	No	Yes	6
No	Yes	Yes	1
Yes	No	No	20
No	Yes	No	9
No	No	Yes	1
No	No	No	5
46	**30**	**9**	**62**

Note: [a]This table summarizes evidence from the 62 case studies of projects where Odebrecht paid bribes. We consider three corruption mechanisms: tailoring of the bidding process, favorable renegotiations, or ex post extortion. [b] We classified the case studies using legal documentary evidence and investigative media sources. We have documentary evidence on the quid pro quo for 40 projects. We use information from investigative media for the remaining 22 projects. [c]In 5 of the 62 projects, we do not have enough information to determine the corruption mechanism and the quid pro quos.

of IIRSA Norte highway, the Peruvian section of an East-West transcontinental highway. Odebrecht agreed with the head of the unit in charge of public-private partnerships to add $28.3 million in expenses and additional investments. Peruvian law required that the agreement be approved by an arbitration panel. According to the prosecutors, two panel members were paid $110,000 by Odebrecht to ensure that the firm would win the arbitration process.[4]

Multiple Quid Pro Quos

Often Odebrecht bribed officials and politicians at different stages of a project, involving different *quid pro quos*. To see this, in Table 4 we tabulate reasons why Odebrecht paid bribes. In the eight countries for which we have complete data, we found judicial documentary evidence of the *quid pro quos* associated with bribe payments in 45 out of 88 projects. For 17 projects, for which we could find no judicial information, we use data culled from investigative press reports.

We found the following: First, for 46 of the 62 projects where we found evidence of a *quid pro quo*, Odebrecht bribed to manipulate subjective bid criteria to either exclude or disadvantage rivals. Second, in 30 projects, Odebrecht bribed to obtain better terms when renegotiating the contract after the projects were awarded. Third, in nine projects, Odebrecht paid a bribe because a public official threatened to block the project. Extortion has been mentioned in the literature as a reason to pay bribes, but it is less frequent in the Odebrecht case. For 27 out of the 62 projects for which we found evidence of the *quid pro quo*, more than one of these reasons applied.

[4]See "Prosecutor's Office accusation against Odebrecht," Disposición N°10, April 17, 2018, p. 5. Available at https://es.scribd.com/document/378437881/Investigacion-fiscal-contra-los-a-rbitros-caso-Lava-Jato.

The following cases illustrate that Odebrecht bribed different individuals as projects progressed. Consider the Ruta del Sol, a 528-kilometer highway running from Puerto Salgar to San Roque in Colombia. A consortium headed by Odebrecht was awarded a public-private partnership contract in 2010 to build and operate it.

Odebrecht paid $6.5 million to Gabriel Garcia Morales, Vice Minister for Transportation, who ensured that the National Infrastructure Agency (ANI), which tendered the project, tailored the auction to favor Odebrecht. To this effect, it included a discretionary pass/fail qualification stage that verified a bidder's financial capacity, the fulfillment of legal requirements, and the bidder's experience delivering public-private partnerships. As a result of the efforts of García Morales, one of Odebrecht's rivals failed on the experience requirement, and the remaining bidder failed on all criteria. Odebrecht bid close to the maximum that bidders could charge because it expected to be the only bidder in the auction.

Next came bribes to get better terms in renegotiations. Odebrecht paid $4.6 million to Otto Bula, a former congressman, to lobby and bribe government officials and politicians. When asked about whether adding the new road to the original project was admissible, the Colombian National Comptroller replied (our translation): "If the object of a concession contract is to build, maintain, and operate a highway between points A and B, it is clear that any facility not included within that highway, such as an extension to a geographic point C, cannot be agreed upon nor executed as an addition to the original contract." Nonetheless, after being bribed, Luis Fernando Andrade, the head of ANI, added the Ocaña-Gamarra project to the Ruta del Sol concession without an open tender. Bula also bribed a member of the Senate Budget Commission responsible for approving the contract renegotiation. Furthermore, Odebrecht modified the original contract by adding toll plazas and by increasing tolls by 15 percent. Overall, the contract was renegotiated ten times, new works were added, and the total cost increased by 29 percent to $1.25 billion, and completion of the project was delayed by five years until at least 2022.

Consider next the second stretch of the 300-kilometer Southern Interoceanic Highway (IIRSA Sur, section II) in Peru. In 2005, a consortium led by Odebrecht was awarded a 25-year contract to build and operate the highway. This project, which was budgeted at $263 million, was politically motivated as it became apparent early on that the highway would not carry much traffic. To exempt the project from a cost-benefit evaluation, the government tendered it as a Design-Build-Operate public-private partnership and not as a public work. The Public-Private Partnership law left the design of the project to the firm, and bidding for the project took place without a preliminary design, which fast-tracked the adjudication.

At this stage, President Toledo's security chief approached Odebrecht offering to use the president's clout to influence ProInversión, the agency in charge of tendering public-private partnerships in Peru, and ensure that Odebrecht would win the contract. They agreed on a $35 million bribe. However, President Toledo failed to deliver, as he could not get ProInversión to raise the unrealistically low official reference value of the project. This was a problem for Odebrecht, because

as mentioned before, bids for the project could not exceed the reference value by more than 10 percent. Toledo was also unable to deliver on other petitions to modify the tender documents. According to Jorge Barata, President Toledo ended up receiving only a $20 million bribe for helping Odebrecht to win the project.

Odebrecht was the only bidder, and the adjudication was rushed through by ProInversión. After the tender, the contract was renegotiated eight times to add major new works, all without a competitive tender. The cost of the project tripled to $654 million. Then Odebrecht paid bribes to the arbitration judges who adjudicated contractual disputes with the Peruvian state. For example, Horacio Cánepa received $1.4 million both to vote in favor of Odebrecht and to suggest which other judges to bribe. Odebrecht won 10 of 13 arbitration cases.[5]

The last case involves construction of the first line of a Metro system in Lima, Peru. Construction had begun in the 1980s during the first government of President Alan García but remained incomplete until revived in 2006 during García's second presidential term. The government tried to auction the first section of the Línea 1 as a public-private partnership in 2006 and 2008, but there were no bidders. In 2009, the government decided to tender the project as a conventional public work with only a preliminary design. Bids were evaluated using a scoring function that put a 70 percent weight on the technical score and the remainder on the cost bid. One of the components of the technical score was a subjective assessment of improvements to the preliminary design.

In his plea agreement, Odebrecht executive Jorge Barata described the *quid pro quo*. The Vice Minister of Transport and Communications Jorge Cuba offered to tailor the technical requirements in exchange for a $1.4 million bribe. Two officials in charge of scoring the technical proposals connived with Cuba, ensuring that Odebrecht obtained the highest technical score. Five bidders were prequalified, but two were excluded for not exhibiting the required legal documents at the prequalification stage. An additional bidder was excluded at the tendering stage for failure to achieve the minimum requirements. Therefore, only two participants made it to the bidding stage. As mentioned before, the auction rules required bids to lie between 90 and 110 percent of the reference value published by the regulator. Odebrecht obtained the maximum technical score and submitted the minimum cost bid allowed. The second bidder got a slighter lower technical score (99.25 instead of Odebrecht's 100) and its cost bid was 6 percent higher than the reference value. As Barata explained: "[G]etting the highest technical score ensured that we [won the project] if we bid the minimum allowed."

Two years later, the second section of Línea 1 was put to tender. This time, however, Jorge Cuba asked for $6.7 million, and Edwin Luyo, who oversaw the tendering process, received $0.5 million. Barata stated that, had Odebrecht refused to pay the larger bribes, Cuba would have allocated the project to a different

---

[5] See "Prosecutor's Office accusation against Odebrecht Disposicion N°10," Public Prosecutors' office from Perú, April 17, 2018. Available at https://www.scribd.com/document/378437881/Investigacion-fiscal-contra-los-a-rbitros-caso-Lava-Jato.

bidder—which suggests that the bidders competed in bribes. As in the first section of the Línea 1, only two bidders made it to the bidding stage, because other consortia were disqualified on technical grounds. Odebrecht again obtained the maximum technical score, and both consortia submitted cost bids equal to the minimum allowed value. Thus, paying a bribe to receive the maximum technical score was essential for Odebrecht to win this project.

Lima's metro lines were renegotiated several times. The cost of the first section increased by 25.2 percent, while the cost of the second section increased by 47.6 percent. An interesting feature of the contract was a built-in renegotiation clause, which was added by decree after the project was awarded to Odebrecht. The decree allowed Odebrecht to unilaterally increase "unit prices"—the values for the various construction components that are required to build the project—once it had completed the design.

The Size of Odebrecht's Bribes and Profits

The broader research literature suggests that the bribes paid to public officials and politicians are often large. For example, Kenny (2009) concludes that bribes in the infrastructure sector are between 5 and 20 percent of construction costs while Glaeser (2019) reports that highway cost overruns due to corruption lie between 20 and 30 percent of project cost. Olken (2007) measured the difference between what an Indonesian village government spent on a road and a cost estimate by expert engineers. Unaccounted expenditures averaged approximately one-fourth of the total cost of the road. Collier, Kirchberger, and Söderbom (2016) showed that the unit cost of roads is 15 percent higher in countries where corruption, as measured by the World Governance Indicators, is above the median. In 2004 the American Society of Civil Engineers claimed that corruption accounts for an estimated $340 billion of worldwide construction costs each year, around 10 percent of the global construction industry value added of $3.2 trillion (ParentAdvocates. org 2004).[6]

In the Odebrecht case, however, bribes as well as the profits derived from corruption were small relative to the size of the projects. To estimate the size of bribes relative to costs, we obtained data on the final cost, including renegotiation,

[6]There are also estimates of the size of bribes in noninfrastructure projects, with a large variation in the relative size of bribes, ranging from a few percentage points in Iraq's Oil for Food Program (Hsieh and Moretti, 2006) to 80 percent in the primary education program in Uganda (Svensson 2003). Reinikka and Svensson (2004) examine a public education program that offered a per-student grant to cover nonwage expenditures in primary schools. Between 1991 and 1995, schools received only 13 percent of what the central government spent on the program. Olken (2006, 2007) shows that in a large antipoverty program in Indonesia, 18 percent of subsidized rice was stolen and that 29 percent of funds allocated to a road building project disappeared. Di Tella and Schargrodsky (2003) compare prices paid for basic homogeneous inputs at public hospitals in the city of Buenos Aires. They show that prices paid fell by 15 percent during the first nine months after a crackdown on corruption in 1996 and 1997. Kaufmann (2005) and IMF (2016) estimate worldwide bribe payments at roughly 2 percent of GDP.

Table 5
Odebrecht Bribes and Associated Profits, Relative to Investment

	Number	Bribes/ final cost	Profits from bribes/ final cost
All projects	88	0.51%	1.26%
Projects with bribes (legal or media sources)	62	0.79%	1.95%
Projects with bribes (legal sources)	45	0.98%	2.41%

Source: Authors' calculations using data in Table 2.

for the 88 projects. As Table 5 shows, bribe payments as a fraction of final cost were less than 1 percent.[7] Profits due to bribes are larger than bribes but still small at approximately 2 percent of final cost.

Odebrecht's profits from bribing, as a percentage of cost, were low. Nonetheless, they represented a large fraction of Odebrecht's total profits. To see this, we compare the profits made from bribing, as reported by the US Department of Justice, with the overall profits of Odebrecht, as reported in financial statements.[8] These show that profits from all operations between 2004 and 2014 were $2.4 billion on revenues of $286.8 billion, or 1 percent of revenues.[9] In comparison, the US Department of Justice estimated that net profits from bribes were around $2.4 billion. This suggests that most of the profits Odebrecht made during the period were due to bribing.

Our conclusion that almost all profits that Odebrecht made came from bribing assumes that the financial statements measure Odebrecht's profits accurately, and there is little, if any, "tunneling"—that is, no unaccounted for transfers of wealth to the owners and managers. A first independent check on this assumption is to compare Odebrecht's profits to the net worth of the Odebrecht family who owns the firm. Data from *Forbes* suggests that during the period, the net worth of the Odebrecht family stayed in the range of $4–6 billion (as reported by Antunes, 2012, 2013, 2014). A second piece of evidence suggesting that our estimate of profits is in the right ballpark is the size of the fine paid by Odebrecht.

---

[7]In the cases involving Colombia, Ecuador, Panama, and Peru, the respective National Attorneys presented evidence suggesting that bribes paid by Odebrecht were larger than those stated in the plea agreement between Odebrecht and the US Department of Justice. Nonetheless, additional bribes do not change the fact that the total remains relatively small.

[8]Odebrecht is a family-owned firm, so it had no legal requirement to produce and publish audited financial reports. However, we were able to track the information from the annual reports that allowed us to reconstruct sales and profits for every year in 2004–2018, with the exception of 2008. There is no publicly available data for Odebrecht's profits in 2008. During the period Odebrecht issued bonds in the international market, including New York, which requires going through the standard due diligence process.

[9]Interestingly, the low profits as a share of sales nonetheless represent a reasonable return on equity. In 2014, for example, Leahy, Rathbone, and Schipani (2015) report: "Like many construction companies, the emphasis is on volume and keeping costs low, which explains Odebrecht's wafer-thin margins: in 2014, net profits were just $210m from $41bn of sales." However, our own calculations show that this is a reasonable 11.3 percent rate of return on equity in the period ending in 2014.

Initially, the US Department of Justice sought to impose a $4.5 billion fine, but Odebrecht successfully argued that such a fine would lead to its bankruptcy and ultimately paid $2.6 billion, a number close to the $2.4 billion in corporate profits mentioned above.

Odebrecht's revenues and market share increased dramatically following the creation of the Division of Structured Operations in 2006. Odebrecht's construction sales increased from around $2 billion in 2003 to approximately $17 billion in 2016. According to the trade publication *Engineering News Record,* in 2003 Odebrecht was the 31st-largest construction company in the world. In 2016, when Odebrecht signed its plea agreement with the US Department of Justice, it had become the sixth-largest construction company in the world.

Our calculations show that Odebrecht's profitability remained low during the entire period we studied (2004–2014). This is surprising, as it may seem that the frequency of contract renegotiations, plus the ability to bias procurement auctions in its favor, should have allowed Odebrecht to obtain a high profit rate on projects. Certainly, Odebrecht's sales and market share grew quickly following the creation of the DSO in 2006. Nevertheless, its profits as a percentage of sales fell. This seems at odds with Shleifer and Vishny (1993), one of the few papers that posits a model that describes the magnitude of bribes. They point out that a corrupt public official with power to exclude firms can increase the corruption rent (in the same way that a monopolist creates a rent by restricting output). In that case, the firm's profit margin and the bribe should be large. Nevertheless, as we have seen, profits were on the order of 2 percent of the final cost of the projects (see Table 5), and bribes were even smaller. Small profits also suggest that Odebrecht won only a small advantage over other competitors by paying bribes. This poses the challenge of explaining why Odebrecht increased its market share dramatically while its overall profits remained flat.

One possible rationalization is that Odebrecht's CEO engaged in empire building by increasing sales at the expense of profits (Jensen and Meckling 1976). This argument is not very compelling because Odebrecht is a family-owned firm managed by the principal. A related explanation is that Odebrecht increased its market share in the expectation of future increased profitability.

Campos et al. (2019) present an alternative explanation. Say that competition in the initial bids forced Odebrecht to lowball and bid below the anticipated cost of the project in the expectation of making good its losses during the renegotiation stage. Because the creation of the DOS gave Odebrecht an advantage in bribing, in a situation where competition among firms is intense and construction firms have similar costs, that small advantage in bribing will lead to a large increase in market share but not to a large increase in profits. This is similar to the reasoning showing that under Bertrand competition, a small cost advantage increases market share dramatically without a substantial increase in profits. Consistent with this explanation, Odebrecht's advantage was bound to be replicated and four years later, a competing Brazilian construction firm, OAS, created its own bribing unit (*IDL-Reporteros* 2019). This unit was smaller than Odebrecht's, which enjoyed a first-mover advantage.

Some Policy Implications

The Odebrecht scandal improves our understanding of corruption in public infrastructure projects and suggests some possible anti-corruption reforms. A rather surprising observation is that in many countries, even those affected by corruption, auctions of large infrastructure projects were fairly competitive at the bidding stage. Despite Odebrecht being the largest corruption case ever prosecuted under the US Foreign Corrupt Practices Act in its 40-year history, margins and profits were small relative to the size of the projects and so were the bribes Odebrecht paid. Small profits suggest that Odebrecht competed, and small bribes suggest that public officials were unable to obtain large rents by selling access to projects. Indeed, no single agent seems to have been in control of an entire project.

The combination of competition and some transparency at the tendering stage limits the discretion of public officials and may reduce the value of bribing. Indeed, as Jorge Barata, one of Odebrecht's Peruvian executives, revealed in the plea agreement with the US Department of Justice, Odebrecht reduced President Toledo's bribe for not being able to make ProInversión—a technical agency—change the tendering documents. Similarly, improved disclosure of financial information for firms that operate in international bond markets limits their ability to generate funds to pay bribes. Last, a complementary explanation for small bribes is the desire to keep a low probability of detection.

Competitive tendering and transparency requirements at the bidding stage are not a coincidence but rather the result of decades of insistence by multilaterals and academics that governments should procure infrastructure in open and competitive auctions. For example, the World Bank has promoted competitive bidding for the projects it finances since the 1990s. More generally, competitive bidding and some transparency in public auctions for infrastructure have become fairly common in many developing countries, especially in Latin America. As Knack, Biletska, and Kacker (2017) find in a sample of 88 countries, more transparency in the procurement process fosters participation in auctions because firms pay smaller bribes and less frequently. Also, Tran (2009) provides evidence that competitive bidding based on objective criteria such as lowest price reduces corruption, while competitive bidding based on subjective criteria does not.

Nevertheless, it is also clear from the Odebrecht case that there is urgent need for improvements. One obvious reform is to use only objective criteria in tenders to adjudicate infrastructure projects. Governments would rely less on subjective criteria if they would tender well-designed projects. On the contrary, several projects where Odebrecht paid bribes were tendered with only preliminary designs, and part of the subjective evaluations consisted of scoring the improvements proposed by bidders. We conjecture that public officials may include subjective criteria and tender projects with incomplete designs to increase the opportunities for corruption.

In contrast to the bidding stage, there is ample need for improvements in the post-tender stage. Multilaterals have had little success in dealing with contract renegotiations, despite the fact that the monograph establishing that contract

renegotiations of public-private partnerships were pervasive in Latin America (that is, Guasch 2004) was written by an economist at the World Bank. A straightforward improvement is to disclose the information on contract renegotiations and make it easily available to the public, which is seldom done, even in developed countries. A more ambitious reform is to reduce the incentives and ability to bias the renegotiation process. Renegotiations should be subject to review by an independent panel of experts, and additional works should be tendered in open auctions that exclude the firm that won the initial contract. This would increase the government's bargaining power and reduce the rents from renegotiation. Doing so would lower the value of renegotiations and moderate the incentive to pay bribes. Chile's Public-Private Partnership Act of 2010 introduced the above-mentioned reforms, and they were followed by a reduction in renegotiations of more than 90 percent (Engel, Fischer, and Galetovic 2020, see Table 5).

References

Antunes, Anderson. 2012. "The Richest People in Brazil." *Forbes*, August 10. https://www.forbes.com/sites/andersonantunes/2012/08/10/the-richest-people-in-brazil/#26d68d666a26.

Antunes, Anderson. 2013. "The Richest People in Brazil 2013: The Full List." *Forbes*, September 9. https://www.forbes.com/sites/andersonantunes/2013/09/09/the-richest-people-in-brazil-2013-the-full-list/#77f6e9897f1d.

Antunes, Anderson. 2014. "The Complete List of The 150 Richest People In Brazil." *Forbes*, September 18. https://www.forbes.com/sites/andersonantunes/2014/09/18/

the-complete-list-with-the-150-richest-people-in-brazil/#2b06190f33da.

Bajari, Patrick, Stephanie Houghton, and Steven Tadelis. 2014. "Bidding for Incomplete Contracts: An Empirical Analysis of Adaptation Costs." *American Economic Review* 104 (4): 1288–1319.

Burguet, Roberto, and Yeon-Koo Che. 2004. "Competitive Procurement with Corruption." *The RAND Journal of Economics* 35 (1): 50–68.

Burguet, Roberto, and Martin K. Perry. 2007. "Bribery and Favoritism by Auctioneers in Sealed-Bid Auctions." *The B.E. Journal of Theoretical Economics* 7 (1).

Campos, Nicolás, Eduardo Engel, Ronald Fischer, and Alexander Galetovic. 2019. "Renegotiations and Corruption in Infrastructure: The Odebrecht Case." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3447631.

Collier, Paul, Martina Kirchberger, and Måns Söderbom. 2016. "The Cost of Road Infrastructure in Low- and Middle-Income Countries." *The World Bank Economic Review* 30 (3): 522–48.

De Michele, Roberto, Joan Prats, and Isaías Losada. 2018. "Efectos de la corrupción en los contratos de asociaciones público-privadas: Consecuencias de un enfoque de tolerancia cero." Inter-American Development Bank Discussion Paper 00625.

Di Tella, Rafael, and Ernesto Schargrodsky. 2003. "The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires." *Journal of Law and Economics* 46 (1): 269–92.

Engel, Eduardo, Ronald D. Fischer, and Alexander Galetovic. 2020. "When and How to Use Public-Private Partnerships: Lessons from the International Experience." In *Economic Analysis and Infrastructure Investment,* edited by Edward Glaeser and James Poterba. Chicago: University of Chicago Press.

Engineering News-Record Magazine. 2009. "Engineering News-Record Top 225 International Contractors." https://www.enr.com/toplists/2020-Top-250-International-Contractors-Preview

Glaeser, Edward L. 2019. "Incentives, Infrastructure and Institutions: Building, Funding and Managing Public Works the Infrastructure Research Agenda." Presentation at The Finance of Infrastructure Symposium, Hoover Institution, Stanford University, May 17.

Guasch, J. Luis. 2004. *Granting and Renegotiating Infrastructure Concessions: Doing it Right.* Washington, DC: The World Bank.

Guasch, J. Luis, and Stéphane Straub. 2009. "Corruption and Concession Renegotiations.: Evidence from the Water and Transport Sectors in Latin America." *Utilities Policy* 17 (2): 185–90.

Hsieh, Chang-Tai, and Enrico Moretti. 2006. "Did Iraq Cheat the United Nations? Underpricing, Bribes, and the Oil for Food Program." *The Quarterly Journal of Economics* 121 (4): 1211–48.

Huang, Yangguang, and Jijun Xia. 2019. "Procurement Auctions under Quality Manipulation Corruption." *European Economic Review* 111: 380–99.

IDL-Reporteros. 2019. "Cómo funcionó el departamento de sobornos de OAS." February 28, *IDL-Reporteros.* https://www.idl-reporteros.pe/departamento-de-sobornos-de-oas-peru.

International Monetary Fund (IMF). 2016. *Corruption: Costs and Mitigating Strategies.* Washington, DC: IMF.

International Monetary Fund (IMF). 2018. *Peru: 2018 Article IV Consultation-Press Release; Staff Report; and Statement by the Executive Director for Peru.* Washington, DC: IMF.

Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics* 3 (4): 305–60.

Kaufmann, Daniel. 2005. "Myths and Realities of Governance and Corruption." https://papers.ssrn.com/sol3/papers.cfm?abstract_id=829244.

Kenny, Charles. 2009. "Measuring Corruption in Infrastructure: Evidence from Transition and Developing Countries." *The Journal of Development Studies* 45 (3): 314–32.

Knack, Stephen, Nataliya Biletska, and Kanishka Kacker. 2017. "Deterring Kickbacks and Encouraging Entry in Public Procurement Markets: Evidence from Firm Surveys in 88 Developing Countries." World Bank Policy Research Working Paper 8078.

Leahy, Joe, John Paul Rathbone, and Andres Schipani. 2015. "Arrest Brings Odebrecht into the Spotlight." *Financial Times,* July 29. https://www.ft.com/content/26f28250-34f3-11e5-b05b-b01debd57852.

OECD. 2019. *Resolving Foreign Bribery Cases with Non-Trial Resolutions: Settlements and Non-Trial Agreements by Parties to the Anti-Bribery Convention.* Paris: OECD Publishing.

Olken, Benjamin. 2006. "Corruption and the Costs of Redistribution: Micro Evidence from Indonesia." *Journal of Public Economics* 90 (4–5): 853–70.

Olken, Benjamin. 2007. "Monitoring Corruption: Evidence from a Field Experiment in Indonesia." *Journal of Political Economy* 115 (2): 200–49.

ParentAdvocates.org. 2004. "Civil Engineers Call for Curb in Worldwide Corruption." http://

parentadvocates.org/index.cfm?fuseaction=article&articleID=4499.

Pereira, Renée. 2017. "Operacão Lava Jato deixa mais de R $90 bi em obras paradas." *Estadão,* June 18. https://economia.estadao.com.br/noticias/geral,operacao-lava-jato-deixa-mais-de-r-90-bi-em-obras-paradas,70001846435.

Poder Judicial de la Republica Dominicana. 2018. "Public Prosecutor's Office Accusation against Odebrecht." https://pciudadana.org/wp-content/uploads/2019/04/Acusacion-del-MP-de-caso-Odebrecht.pdf.

Poder Judicial del Perú. 2019. "Expediente Nº 00075-2017-3-5001-JR-PE-03." https://www.pj.gob.pe/wps/wcm/connect/85dd6000416cd024a408ad0464bd7500/1C2%B0+SPAN_EXP.+75-2017-3_Revocan+Prisi%C3%B3n+preventiva+F%C3%89LIX+MORENO-reducido.pdf?MOD=AJPERES&CACHEID=85dd6000416cd024a408ad0464bd7500.

Pressly, Lisa. 2018. "'The Largest Foreign Bribery Case in History." *BBC News*, April 21. https://www.bbc.com/news/business-43825294.

Reinikka, Ritva, and Jakob Svensson. 2004. "Local Capture: Evidence from a Central Government Transfer Program in Uganda." *The Quarterly Journal of Economics* 119 (2): 679–705.

Shield, Fergus, and Sasha Chavkin. 2019. "Bribery Division: What Is Odebrecht? Who Is Involved?" *International Consortium of Investigative Journalists*, June 25. https://www.icij.org/investigations/bribery-division/bribery-division-what-is-odebrecht-who-is-involved/.

Shleifer, Andrei, and Robert W. Vishny. 1993. "Corruption." *The Quarterly Journal of Economics* 108 (3): 599–617.

Simon, Roberto. 2019. "Odebrecht: LatAm's Biggest Stories of the 2010s." *America's Quarterly*, December 18. https://www.as-coa.org/articles/odebrecht-latam%E2%80%99s-biggest-stories-2010s.

Smith, Michael, Sabrina Valle, and Blake Schmidt. 2017. "No One Has Ever Made a Corruption Machine Like This One." *Bloomberg Businessweek*, June 8. https://www.bloomberg.com/news/features/2017-06-08/no-one-has-ever-made-a-corruption-machine-like-this-one.

Svensson, Jakob. 2003. "Who Must Pay Bribes and How Much? Evidence from a Cross Section of Firms." *The Quarterly Journal of Economics* 118 (1): 207–30.

Tran, Anh. 2009. "Can Procurement Auctions Reduce Corruption? Evidence from the Internal Records of a Bribe-Paying Firm." Unpublished.

US Department of Justice. 2016. *Odebrecht and Braskem Plead Guilty and Agree to Pay at Least $3.5 Billion in Global Penalties to Resolve Largest Foreign Bribery Case in History*. Washington, DC: US Department of Justice.

US Securities and Exchange Commission. 2008. "SEC Files Settled Foreign Corrupt Practices Act Charges Against Siemens AG for Engaging in Worldwide Bribery With Total Disgorgement and Criminal Fines of Over $1.6 Billion." Litigation Release No. 20829.

The Rise of Research Teams: Benefits and Costs in Economics

Benjamin F. Jones

Economics research is increasingly a team activity: economists increasingly coauthor their papers, and these coauthored papers have a large and increasing impact advantage. This "rise of teams" raises issues for individual researchers and for the field. On the one hand, coauthorship brings benefits, allowing individuals to combine perspectives, knowledge, skills, and effort in fruitful ways. But it also imposes costs; for example, coauthorship divides and obscures credit among the participants, which can undermine individual career progression. This paper synthesizes recent literature to weigh the benefits and costs of research teams. The findings provide guidance to individual researchers themselves, and the institutions that support them, in fostering high-impact research and productive research careers.

The paper begins by documenting the rapid rise of team authorship in economics. For example, while papers with two or more authors constituted only 19 percent of economics journal articles in 1960, this share rose to 44 percent in 2000 and 74 percent in 2018. Moreover, team-authored papers in economics have increasing impact advantages over solo-authored papers. By 2010, a team was three times more likely to produce a highly cited paper than a solo author, an advantage that has grown steadily with time. These shifts appear not only within every subfield of economics, but also in virtually all fields of science, social science, and patenting.

■ *Benjamin F. Jones is Gordon and Llura Gund Family Professor of Entrepreneurship and Professor of Strategy, Kellogg School of Management at Northwestern University, Evanston, Illinois, and Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His email address is bjones@kellogg.northwestern.edu.*

This shift to teamwork brings both advantages and problems. The multi-disciplinary literature on creativity emphasizes that creativity is a process of combination—a novel mixture of existing material and methods (for example, Schumpeter 1939). Viewed as "combinations of ideas," high-impact research turns out to combine prior knowledge in distinctive ways. Moreover, teams are more likely to produce these distinctive creative combinations. For example, teams appear to navigate extant knowledge to produce more novel combinations of ideas (Uzzi et al. 2013).

Meanwhile, one individual can only know so much. As scientific knowledge accumulates with time, individuals appear restricted into ever-narrower subspecialties of expertise. This increasing specialization can help explain the rising relative advantage of teams or, put another way, the declining impact of solo researchers (Wuchty, Jones, and Uzzi 2007; Jones 2009). Beyond aggregating differentiated knowledge, however, teams may also reflect vertical differences in the productivity of the coauthors. These differences may in turn have performance implications. For example, one might imagine that the quality of team output follows from the strongest member of the team, who might be the creative engine or otherwise drive the enterprise. On the other hand, team output may follow from the weaker members of the team, perhaps due to bottlenecks at certain tasks. Perhaps surprisingly, team impact in economics, as in all other fields, is weighted toward the lower-impact rather than higher-impact team members (Ahmadpoor and Jones 2019). Consistent with this finding, scholars engage in positive assortative matching when forming teams, in all fields.

Overall, striking empirical regularities emerge when considering the rise of teams. Echoing how Zipf's Law describes the size distribution of cities (Gabaix 1999), or the Kaldor (1961) facts discipline macroeconomics, a substantial surprise of the innovation literature is that the production of ideas—which might seem to be a messy and opaque creative exercise that defies ready description—is given to strong empirical regularities.

These same regularities, however, also point to particular costs. For one, teamwork obscures credit. Teamwork can thus put stress on the reward system of science, where tenure and promotion, prizes, and status more generally all depend on the community's assessment of individual scholars. Teamwork seems to undermine this system, at least in the community's capacity to rely on objective indicators. For example, consider the career implications now that almost all work, and an even greater share of the high impact work, is coauthored. When work is coauthored, each paper provides less of a signal about the individual authors. Yet tenure clocks have not lengthened. Should lifetime tenure contracts be awarded based on a short series of increasingly intertwined signals? Moreover, economics (like many fields) continues to award prizes to individuals rather than teams. But do early-career prizes like the John Bates Clark medal, which is awarded by the American Economic Association to a prominent American economist under the age of 40 (described at https://www.aeaweb.org/about-aea/honors-awards/bates-clark), which have traditionally celebrated individuals, make sense when more and more work—and the highest-impact work—is done in teams? If economics is increasingly relying on

subjective assessments of credit within a group effort, given the rise of teams, do tenure committees, funding panels, and prize committees increasingly impose a series of biases related to gender, personal relationships, institutional eminence, or other features that lead to discrimination? These concerns are not hypothetical: recent work in economics suggests that women, unlike men, are less likely to receive tenure when their work is coauthored (Sarsons et al. 2021). While economics is shifting sharply toward teamwork, it's far from clear that the institutions that support research are keeping up with the rise of teams.

The Rise of Teams

A large literature has studied "team science" through the perspective of coauthorship (for example, Adams et al. 2005; Wuchty, Jones, and Uzzi 2007). Although teamwork and collaboration in research can extend beyond formal coauthorship, benefits of a coauthorship orientation include its ease of measurability with large databases and its direct relevance to career progression. Authorship records form central measures of scientific careers, as in academic curricula vitae, and are a primary basis for community evaluations of scholars, including in tenure decisions, as will be discussed further below. This section charts the rise of teams in economics through the lens of coauthorship.[1]

The Rising Frequency of Teams

Publications in economics were once largely a solo author's game. Examining economics papers published before 1900, about 90 percent were solo-authored. This pattern holds true in my own calculations using the 1.7 million economics papers indexed from 1816–2018 in the Microsoft Academic Graph (data described in Sinha et al. 2015), and similar patterns appear using the Web of Science™ data published by Clarivate (Wuchty, Jones, and Uzzi 2007). The fraction of solo-authored work declined in the first half of the 20th century, but only modestly, when about 80 percent of economics papers were solo-authored. The dominance of solo-authorship was even starker looking among today's top five journals—following common practice, these are defined here as the *American Economic Review, Econometrica, Journal of Political Economy, Review of Economic Studies,* and *Quarterly Journal of Economics*—where 98 percent of articles were solo-authored before 1950.[2]

But then the pattern began to change. Figure 1, panel A presents some trends. Solo authorship represented 80 percent of economics papers in 1960 and 65 percent

---

[1] Research on "team science" further embraces co-invention in technology, including patenting and software development (for example, Wu et al. 2019). Alternative constructs of teamwork and collaboration in the sciences can extend from non-coauthor research assistants to those who provide comments and advice, where the "invisible college" of science suggests potentially open collaborative boundaries (Oettl 2012).

[2] Perceptions of the top journals evolve over time. This set of journals is meant as one benchmark using prominent journals.

Figure 1
The Rise of Team Size in Economics

Panel A. All economics papers, 1950–2018

Panel B. Top-five economics journals, 1950–2018
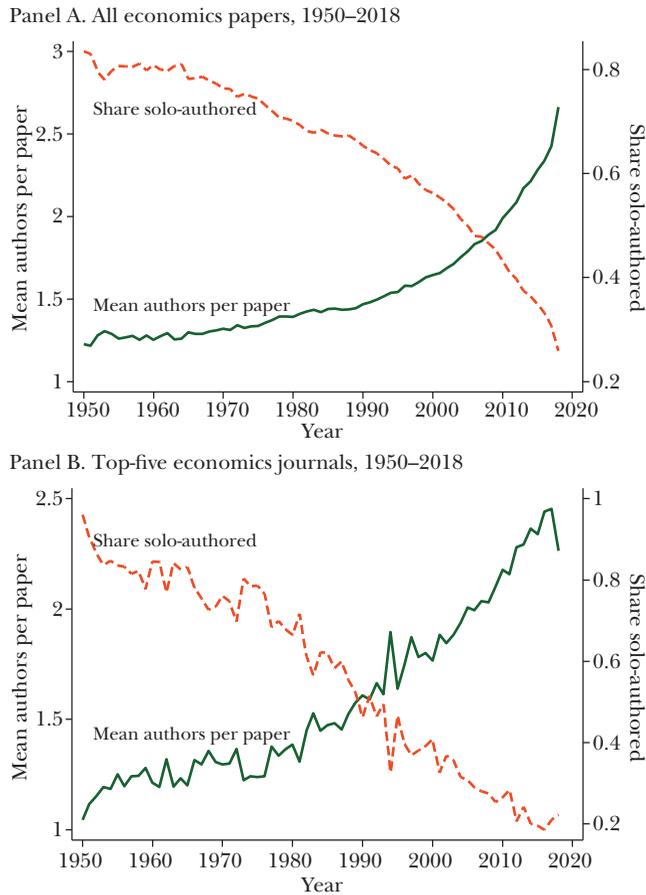
Note: Data are from Microsoft Academic Graph, which indexes 1.7 million papers in economics from 1816–2019. For details, see the online Appendix available with this article at the *JEP* website.

in 1990, but then solo-authorship fell out of the majority in 2005 and represents only 26 percent of economics papers today (as measured by the right-hand axis). To put it another way, in 1950, there were 1.2 authors per economics paper. Average team size reached 2.0 for the first time in 2010. By 2018, team size averaged 2.7 (as shown on the left-hand axis). The jump in average team size in economics papers over the last ten years is greater than the jump over the prior half-century.

Figure 1, panel B shows the trends for the top five economics journals. Here we see similar patterns, albeit with a more linear dynamic. In 1950, solo author-ship was more common among these journals (96 percent) than among economics

publications overall (83 percent). Now, solo authorship is less common in these top-five journals (22 percent) than in economics overall (26 percent). This reversal leads us toward the second set of facts: the relationship between team-authorship and high-impact work.

The Rising Impact of Teams

As a measure of success, define "home-run" papers in a given year as those in the top N percentile of citations received among all publications that year: thus, home-run papers may be defined as those in the top 10 percent, 5 percent, and 1 percent of citations received. The home-run measure is normalized by year of original publication, so it is not affected by time trends in total number of citations. We can then define the relative team impact (RTI) as

$$RTI = \frac{team\ home\ run\ rate}{solo\ home\ run\ rate}$$

where the numerator is the fraction of team papers that turn out to be home runs, and the denominator is the fraction of solo-authored papers that turn out to be home runs (Wuchty, Jones, and Uzzi 2007).

Figure 2, panel A presents the relative team impact over time. Teams have a growing impact advantage. In addition, this growing advantage is stronger when one looks at higher thresholds of impact. From the 1950s through the 1970s, a team-authored paper was 1.5 to 1.7 times more likely to become a home-run than a solo-authored paper, with the modest variation depending on the impact threshold. By 2010, the home-run rate for team-authorship was at least 3.0 times larger than for solo-authorship. From the 1980s onward, the team-impact advantage is increasing as the impact threshold rises. By 2010, team-authored papers are 3.0 times more likely to reach the top 10 percent of citations, 3.3 times more likely to reach the top 5 percent of citations, and 4.1 times more likely to reach the top 1 percent of citations than solo-authored papers.

The impact advantage of teams appears strongly at even higher impact thresholds, as well (not shown in the figure). The data thins at higher thresholds of impact but also shows a rising advantage of teams. For example, defining home runs at the top 0.1 percent in citations received in a given year, the relative team impact was below 1.0 in the 1950s and 1960s, 1.4 in the 1970s, 2.9 in the 1980s, and over 3.0 in each decade since 1990. This pattern is also evident if one looks at eminent individuals. For example, consider the highest impact work by winners of the Clark medal, utilizing each work's citation impact in Google Scholar. For Gary Becker, who won the award in 1967, all five of his top five and nine of his top ten publications are solo-authored. By contrast, for Clark medalists in the last decade, the median case shows one of the top five and two of the top ten publications being solo-authored.

One might imagine that the team advantage is increasing because teams are getting larger and larger. Perhaps more authors simply mean more citations. To address this possibility, Figure 2, panel B presents the "relative team impact" holding the number of authors fixed. This figure uses the 5 percent definition of home

Figure 2
The Rise of Team Impact in Economics

Panel A. Relative team impact, 1950–2010

Panel B. Relative team impact for different size teams
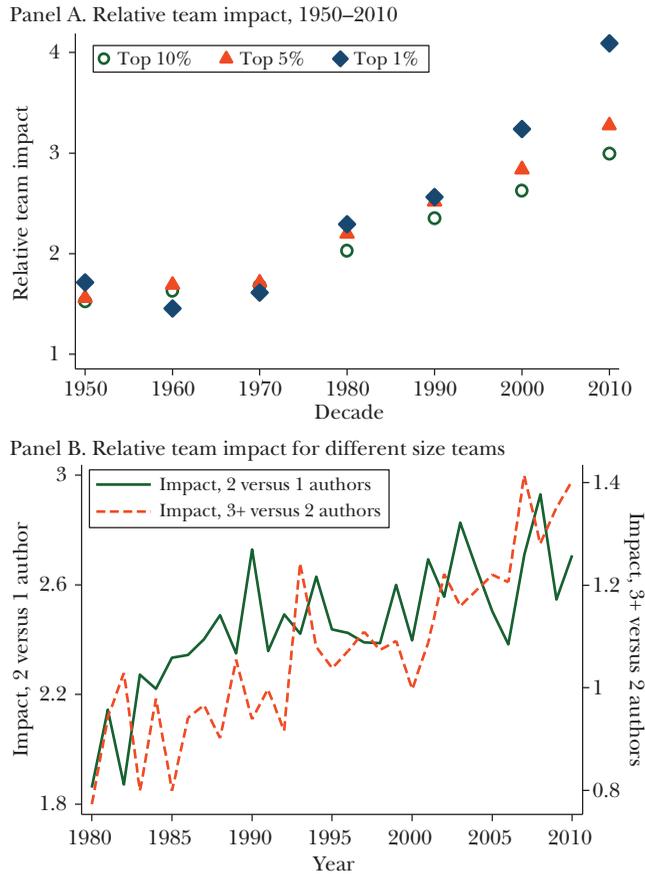
Note: Relative team impact is the percentage of team-authored papers that achieve a given citation threshold divided by the percentage of solo-authored papers that achieve a given citation threshold. For details, see the online Appendix available with this article at the *JEP* website.

run papers, but similar trends appear using different thresholds for defining home runs. In the first comparison (measured on the left axis) we consider two-authored papers versus solo-authored papers, and we see a sharply increasing impact advantage over time. In other words, the upswing is not about adding more authors, but rather a given number of team authors has an increasing impact compared to solo authors.

Figure 2, panel B also compares larger author teams to two-author teams. Three-author teams were substantially less than 5 percent of economics papers prior to 1980, so we focus on post-1980 data to increase sample size. By 2018, three-author teams represent 23 percent of all economics publications. Interestingly, teams with

at least three authors initially underperformed two-author teams: in the 1980s, these larger teams' papers were about 10 percent less likely than two-authored papers to become home runs. This pattern has now reversed. The home-run rate for the larger teams has risen consistently compared to two-authored teams. By 2010, the larger-team papers were 1.4 more likely to be home runs than two-author papers and vastly more likely to be home runs than solo-author papers.

Generality across Research Areas

One might imagine that the rising frequency of teams in economics, or the rising impact advantage of teams, may be confined to a few large subfields. However, these patterns instead appear to be systematic across sub-branches of economics research.

Figure 3, panel A presents the shift toward team authorship, with economics organized into 16 different subfields. For all subfields, team authorship has been more prevalent since 2015 than it was in the 1950s. Four fields—economic history, law and economics, political economy, and development economics—continue to show substantial solo authorship, but nonetheless exhibit increased teamwork with time. Moreover, leading journals within these fields show a greater team orientation today. For example, looking at top field journals in development economics (here taking the *Journal of Development Economics*, *Economic Development and Cultural Change*, and *World Development*), one sees that 75 percent of papers have been team-authored since 2015, which is much more in line with the broader trends. To further explore field level generality, Table 1 narrows the focus to some top field journals for many prominent economic subfields and shows a systematic shift to teams.

Figure 3, panel B presents the relative team impact within each subfield, taking the 1950s as a baseline and comparing the relative team impact for papers published in the 2000s. Specifically, we look at all publications in the years 2000–2009, which gives a large sample for each field and substantial time after the publication year to count citations to each paper. The home-run rate is now measured at the subfield level: specifically, a home run is a paper in the upper 5 percent of citations received among all papers published in the given subfield and year. The relative team impact has gone up over time in every subfield. Notably, for about one-third of subfields, the relative team impact was below one in the 1950s. That is, solo authors substantially outperformed team authors in producing home runs. These were large and central fields: macro, micro, econometrics, labor, and public. However, after 2000, these fields have experienced a reversal, showing substantial team advantages.

The Microsoft Academic Graph data used here does not specifically denote a "theory" subfield within economics. However, one can look at theory-oriented journals. For example, the rise in teams appears strongly when studying *Econometrica*, *Games and Economic Behavior*, and the *Journal of Economic Theory*. Team-authored work in these journals has risen from 39 percent in the 1980s to 71 percent of papers in 2018. The relative team impact measure for these journals averaged 4.1 from 2000 to 2009. Generalizing to some top journals in each field, Table 1 shows the impact advantage of teams is systematic.

Figure 3
Generality across Economic Subfields

Panel A. The rise of team size

Panel B. The rise of team impact
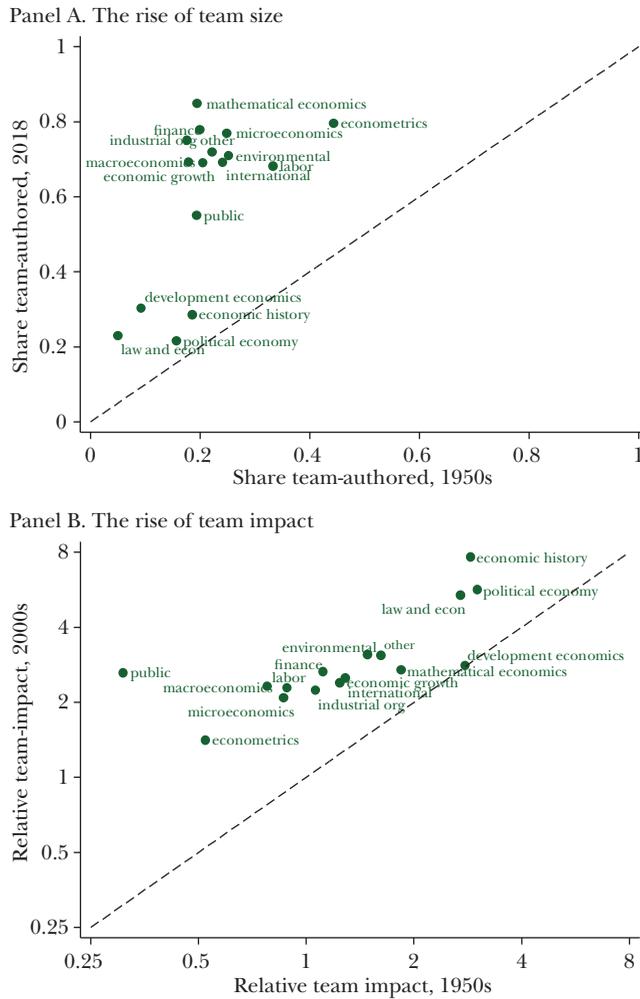
Note: The sixteen different subfields of economics are developed from the field coding in the Microsoft Academic Graph. For details, see the online Appendix available with this article at the *JEP* website. Data points above the 45-degree line indicate a shift toward teamwork (panel A) and rising impact of teams (panel B).

That these results generalize within economics should not be surprising in light of broader literature in the sciences, social sciences, and patenting. Regarding the frequency of collaboration, economics is following in the footsteps of the hard sciences, where the majority of papers were already team-authored in the 1950s (Wuchty, Jones, and Uzzi 2007). But far more broadly, the rising frequency of teams, and the rising impact advantage of teams, extend across virtually all fields of scientific inquiry and

Table 1
Recent Team Prevalence and Impact Advantage, Top Field Journals

Top field journals	Share team-authored, 2018	Share team-authored, 1980s	Relative team impact, 2000s
Development	0.72	0.25	1.92
Econometrics	0.86	0.41	1.63
Finance	0.81	0.55	3.96
Economic geography	0.64	0.50	3.11
Industrial organization	0.78	0.43	1.68
International	0.78	0.27	2.09
Labor	0.75	0.44	1.12
Law and economics	0.80	0.43	4.06
Macroeconomics	0.74	0.34	4.71
Public	0.69	0.41	2.60
Theory	0.71	0.39	4.12

Notes: This table considers the frequency of team-authored papers and the relative impact advantage of teams, studying the top journals in each field. The journals are the top three by field according to current citations metrics (SCImago Journal Rank) and where the Microsoft Academic Graph data has substantial coverage. For details, see the online Appendix.

all technology classes of patenting (Adams et al. 2005; Wuchty, Jones, and Uzzi 2007; Jones 2009). This generality notably extends not only to distant fields throughout science and engineering, but also to "nearer neighbors" to economics, such as political science, sociology, and psychology. For example, comparing the late 1990s to the late 1950s, the fraction of team-authored papers had risen by 270 percent, 90 percent, and 115 percent in political science, sociology, and psychology, respectively, and the relative team impact in these neighboring fields has become as large or larger than in economics (Wuchty, Jones, and Uzzi 2007).

The generality in the "rise of teams" suggests very broad forces are at work. While one can consider mechanisms that may affect particular sub-fields in certain ways, what is happening in economics and across its many sub-fields, is happening across the entire landscape of the social sciences, hard sciences, and engineering as well as in patenting. The universality of these changes, despite different field-level norms and institutions, including the different research settings of universities, government labs, and for-profit businesses, suggests that very general forces are at work. With that generality in mind, we can examine where the rising advantage of teamwork may come from, emphasizing empirical evidence and perspectives that apply across the landscape of research.

The Benefits of Teams: Dimensions of Advantage

Teamwork as Knowledge Aggregation

The rising team advantage can be framed on one dimension as rooted in the accumulation of scientific knowledge. Further research often builds on prior

knowledge—as Isaac Newton said, "[I] if I have seen further, it is by standing on ye sholders of giants" (Inwood 2003). But this progress of science can create a problem for the individual scholar, where the cumulativeness of knowledge can make it increasingly difficult for an individual to be broadly expert across the knowledge frontier (Jones 2009). In Albert Einstein's (1941) words: "[K]nowledge has become vastly more profound in every department of science. But the assimilative power of the human intellect is and remains strictly limited. Hence it was inevitable that the activity of the individual investigator should be confined to a smaller and smaller section."

Einstein's "inevitable" specialization in turn naturally leads to teamwork. As individual researchers become increasingly narrow, teams allow the aggregation of specialized knowledge and thus offer a line of continued attack on problems of wider application (Jones 2009). This force—cumulativeness leading to increasing narrowness—provides one inroad to explaining the increasing tendency to work in teams across all fields and, more particularly, the declining impact of solo authors compared to teams.

To put some empirical content around this conceptual perspective, consider that John Harvard's collection of approximately 400 books was considered a leading collection of his time, and its bequest in 1638, along with small funds for buildings, helped earn him the naming right to Harvard College (Morrison 1936). One hundred seventy-five years later, Thomas Jefferson's renowned library of 6,487 books formed the basis for the US Library of Congress. That library's collection had risen to 55,000 books by 1851 (Cole 1996). Today, the US Library of Congress holds 39 million books (as described in https://www.loc.gov/about/general-information).

Looking instead at journal articles, the flow rate of new papers grows at 3–4 percent per year. In 2018, peer-reviewed, English-language journals published three million new papers (Johnson, Watkinson, and Mabe 2018). In total, the Web of Science™ now indexes 53 million articles from science journals and another 9 million articles from social science journals (as described at https://clarivate.com/webofsciencegroup/solutions/web-of-science). In economics alone, the Microsoft Academic Graph counts 30,100 economic journal articles published in the year 2000. This publication rate was twice what it was in 1982 and half what it is today. The number of high-impact papers has also become very large; for example, among publications in the year 2000 alone, 2,849 economics articles have received at least 100 citations. To Einstein's point, it would seem increasingly difficult for an individual economist to stay on top of the flow of new ideas, or even the flow of relatively impactful ideas, let alone the stock of existing ideas.

Jones (2009) denotes the ensuing challenges as a "burden of knowledge," where individuals respond to cumulativeness along two dimensions. The first is the length of training: individuals can engage in longer training phases, like pre-doctoral programs, lengthening the time of the PhD program, and then post-doctoral programs, to acquire expanding stores of knowledge (Jones 2009, 2010). The second is the increasing narrowness that Einstein described, confining

researchers in the reach of their ideas. This rising narrowness can also be quantified in data. For example, individual researchers are less likely to switch subfields with time, including economics, which is consistent with increased specialization (Jones 2009; Schweitzer and Brendel 2019). Looking at a point in time, researchers are also less likely to switch subfields when in deeper areas of knowledge (Jones 2009).

The organizational implication—teamwork—then follows naturally as a means to aggregate expert knowledge. In the history of aviation, for example, the Wright brothers designed, built, and flew the first heavier-than-air aircraft in 1903. This pair of individuals successfully embraced and advanced extant scientific and engineering knowledge. Today, by contrast, the design and manufacture of airplanes calls on a vast store of accumulated knowledge and engages large teams of specialists; today, 30 different engineering specialties are required to design and produce the aircraft's jet engines alone.[3]

The role of teams in aggregating knowledge appears in diverse empirical contexts. For example, large literatures in psychology emphasize the value of teams in aggregating diverse information to solve problems. A meta-analysis of 72 psychology studies indicates that team performance is strongly increasing when individuals bring distinct information sets and share their information across the group (Mesmer-Magnus and DeChurch 2009). In research teams, survey evidence also links teamwork to specialization. When asked about the primary reasons for collaboration—which could include access to funding, data, physical tools and laboratories, communications advantages, the joy of working together, or specialized skills—by far the dominant answer scientists gave was access to individuals with unique knowledge, expertise, or capabilities (Freeman, Ganguli, and Murciano-Goroff 2015). A separate survey of researchers in 20 fields of science and social science links increased team size with an increasing division of labor and the aggregation of diverse fields of expertise (Lee, Walsh, and Wang 2015). A natural experiment in mathematics also links increased teamwork to exogenous shocks in access to specialized frontier knowledge (Agrawal, Goldfarb, and Teodoridis 2016).

In short, the greater the stock of knowledge in an area, the narrower the expertise of the individual investigator becomes, and the greater the role of teamwork in attacking broad problems. In fact, teams are not only larger with time but also larger at the same point in time when looking at deeper areas of knowledge, which can be measured by the size of the stock of referenced information (Jones 2009). From this perspective, economics can be seen as lagging the hard sciences, where teams are larger and the rise of teams began much longer ago. To the extent that the sciences have accumulated more knowledge historically, scholars in the social sciences may naturally have remained comparatively less specialized and less team-intensive for longer. The rise of teams in economics can then be seen as a sign of the progress of the field. In any case, teams have now come to economics and, as in the sciences, the frequency and impact advantages of research teams only appear to grow.

---

[3]This point is discussed in Jones (2014); Joseph Palladino of General Electric Aircraft Engines provided this specific estimate in personal correspondence.

Teamwork as Creative Combination

Creativity is widely seen as a process of *combination*, where existing material is drawn together in fruitful, new ways (Schumpeter 1939; Fleming 2001). This perspective appears in economics, psychology, sociology, art history, and histories of science of technology among other fields (Usher 1954; Becker 1982; Weitzman 1998; Schilling 2005; Uzzi and Spiro 2005; Rothenberg 2015). This viewpoint also motivates policies and funding mechanisms devoted to interdisciplinary research (National Research Council of the National Academies 2015), which can be construed as a search for advantageous combinations by combining people across disciplinary boundaries.

Recent research has found ways to bring data to bear on this combinations perspective. The research shows that high-impact work combines prior knowledge in distinctive ways. Moreover, teams are more likely to produce these distinctive creative combinations. For example, Uzzi et al. (2013) study the reference lists of 18 million papers. Kim et al. (2016) study the technology code combinations in 8.8 million US patents. These papers define whether any given pairing (of referenced journals, or of technology codes) is relatively "conventional" or "novel." Each new work contains a distribution of such combinations, and two findings emerge: high-impact work is distinctive for 1) the extreme conventionality of most of its combinations, yet 2) the presence of "tail novelty"—a small set of highly unusual combinations. The highest-impact work thus appears simultaneously to be exceptionally heavily grounded in convention, while introducing a truly unusual pairing of prior work. Absent either this depth in conventional combination or the novel combinatorial edge, the chance the paper or patent becomes high-impact falls in half.

Another combinatorial perspective examines mixtures of old and new ideas. Here again there are striking regularities. Mukherjee et al. (2017) study the age profile of references in 28 million papers and five million patents. The highest-impact papers are found to draw 1) exceptionally heavily on recent work yet also 2) relatively widely through time across the corpus of prior knowledge. Absent either element, the chance the new work becomes high-impact again falls in half.

These "combinations of ideas" findings are virtually universal across fields, appearing as general descriptive rules that locate high-impact work. The findings also suggest that the creative combinations problem is not easy. Researchers appear to achieve high-impact when accessing knowledge widely across time and beyond convention—suggesting the creative search problem extends across an enormous landscape of knowledge.

Teamwork, in turn, is strongly associated with these particular creative combinations. Teams are far more likely to achieve the mixture of hyper-recent and older-standing knowledge. Also, team papers are substantially more likely to contain tail novelty. Notably, these combinatorial rules also appear among solo authors. In particular, a solo-authored work that contains these distinctive combinatorial features tends to be much higher impact than solo-authored work that is missing these features. What is distinctive about teamwork, then, is not that teams operate according to different underlying combinatorial rules. Rather,

teamwork is associated with a sharply higher tendency to achieve these fruitful creative combinations. Where tested, these findings also appear net of individual fixed effects, even among notable scientists. For example, a given Fields medalist in mathematics is more likely to achieve the distinctive mixture of old and new, and associated higher impact, when working in a team (Mukherjee et al. 2017).

Finally, teams of different size appear to do different things. Based on citation networks, one can measure the extent to which a paper develops or disrupts prior ideas (Funk and Owen-Smith 2016). Applying these measures systematically in large datasets, smaller teams prove more likely to produce disruptive ideas, while large teams are more likely to develop and consolidate existing ideas (Wu, Wang, and Evans 2019). These measures have been cross-validated in several ways. For example, surveys of scholars across fields have identified disruptive versus developmental papers, and these independent categorization efforts prove highly correlated with the citation-network measure (Wu, Wang, and Evans 2019). These findings, which generalize across papers, patents, and software innovations, as well as when comparing among the works of a given author, suggest a subtler interpretation of the nature of creativity in teams. Namely, disruption—or more revolutionary forms of creativity—appears difficult to manage or produce among large sets of people, and thus appear as the domain of small teams. By contrast, the developmental (and still highly impactful) fruits of "normal science," which advances and refines existing paradigms, appear to be the domain of larger teams, consistent with an ability to aggregate extant knowledge and apply differentiated expertise to advance against known problems.

Teamwork as Vertical Combination

Beyond "horizontal" combinations of people with diverse expertise, recent work also investigates teams as "vertical" combinations, where relatively high-impact and low-impact individuals collaborate. A primary question is whether the joint output reflects the typical output of the higher- or lower-impact team members. At one extreme, team output might follow a "max" process, where the only person who matters is the top person, perhaps because this person generates the creative ideas and/or determines the overall research direction of the group. At the other extreme, the output might follow a "min" process, where the weakest member of the team determines the joint outcome, perhaps because this person creates bottlenecks at certain important tasks.

Ahmadpoor and Jones (2019) examine this question by tracking millions of individual authors and inventors through their collaboration networks. The outcome studied is the citation impact of each individual paper or patent produced. Because the same individual will typically work with different sets of collaborators, and may also occasionally work alone, one can identify an individual fixed effect for each researcher. Simultaneously, one can examine the functional form for the collaborative outcomes for researchers with different fixed effects. The universal finding is that team output is predicted more by the lower-impact members rather than the higher-impact members of the team. This finding appears in all fields of

the sciences and social sciences, including economics as well as in patenting. The joint output typically follows the harmonic or geometric average of the individual fixed effects, which heavily weight lower values when averaging.

Despite this "averaging down" pattern, there is simultaneously a large advantage to teamwork. This advantage appears net of the individual fixed effects. For two people of the same measured quality, their team will typically achieve approximately double the citation impact as when these same individuals work alone. This benefit means that teams, even with some diversity in the vertical quality of the team members, still tend to produce papers (or patents) with more impact working together than working separately.

While it may be surprising that team output tends toward the lower-impact member, these findings are consistent with substantial complementarity in the tasks each team member performs. With teams aggregating diverse expertise through complementary skills or a division of labor (Jones 2009; Freeman et al. 2015; Lee et al. 2015), production functions that emphasize bottlenecks are natural. Conceptually, a top member of the team may still provide creative direction, and elevates the team potential, but to the extent that implementation requires complementary tasks, joint output becomes ultimately and more strongly determined by limited success at specific tasks.

Such within-team complementarities also have organizational implications. Namely, the efficient organizational form then features individuals of similar vertical quality working together (which is referred to as positive assortative matching). Not surprisingly, if perhaps for a variety of reasons, this organizational tendency is also seen in the data, where positive assortative matching is the norm in every field of science and social science, including economics and in patenting (Ahmadpoor and Jones 2019). This is the opposite of what a "max" like function would imply, in which case the efficient organizational form would be to spread the best people around into independent teams.

Teamwork as a Laboratory

Vertical and horizontal components of teamwork can also help inform the "laboratory model" of research, which is common in the hard sciences and, anecdotally, appears to be increasingly common within economics. This research model includes principal investigator(s) as project leader(s) and a variety of tasks executed across a hierarchical team. Field experiments in economics, which can require many researchers to execute, as well as empirical projects that rely on the creation, integration, or heavy computational analysis of large datasets, provide examples within our field. The laboratory model takes on distinct organizational forms and draws on distinct skills. The research team may extend beyond the coauthor list on the ultimate research article, and team leaders typically have substantial overarching control—in designing the project, hiring team members, assigning individual roles, communicating the results, and managing funding. The principle investigator(s) must also monitor the potentially difficult-to-observe execution by each team member and engage in substantial coordination efforts.

In understanding the rise of teams in economics, note that laboratory models in economics still appear quite rare, and the model applies less to many economics subfields, such as theory, that are nonetheless also experiencing the rising frequency and impact advantage of teams. Looking broadly at Nobel prizewinners and Clark medal-winners over the last decade, one is struck by the prevalence of theoretical contributions. Nonetheless, laboratory models do appear increasingly prominent in economics. For example, the 2019 Nobel prize, which was awarded to Abhijit Banerjee, Esther Duflo, and Michael Kremer, summarizes their contributions as being "for their experimental approach to alleviating poverty." Among the last decade's Clark medal-winners, several awardees appear to have leveraged laboratory-like research models, which further hints at the success of the approach. More generally, both field experiments and big data analyses are on the rise, with the former deploying potentially large data-collection teams in the field and the latter often requiring substantial data infrastructure work and often deploying methodological advances, including network methods, text analysis, and machine learning techniques, that have extended researchers' toolkits in computationally-intensive ways (Currie, Kleven, and Zwiers 2020). These approaches can all lend themselves to laboratory models.

As with teamwork in general, the advent of laboratory models in economics suggests another way in which economics is following in the footsteps of the hard sciences. One implication is that funding becomes an increasingly important input for research, as these models are resource-intensive. Funding constraints can then be especially impactful, and these models may already be substantively constrained in the economics field where US federal research expenditure for the economics and the social sciences, both in total and per researcher, remains tiny compared to the hard sciences.[4] Funding constraints may, in turn, raise equity issues, where a small number of elite researchers, or elite institutions, may be privileged with better funding opportunities and seize upon the advantages of high-scale laboratory teams. As we will consider below, the hard sciences may provide useful models for how economics can navigate these issues.

Teamwork and Communications

Finally, the rise of teams may be viewed in light of the advance of information and computing technologies, which has made collaboration easier. This appears especially true for geographically distant collaboration, which has substantially

---

[4]The social sciences as a whole received 1.8–2.9 percent of US annual federal funding since 1990, but social science PhDs are typically 8.6–9.5 percent of US PhD recipients over this period. This indicates that not only is federal funding for social sciences tiny in total, it is even lower compared to the sciences when measured per PhD in the field. Further, the federal research funding share of the social sciences and of economics has been declining with time, even as the PhD shares have been steady. From this perspective, resource-intensive research in economics appears both uncommon and relatively constrained (National Center for Science and Engineering Statistics, Science and Engineering Indicators 2018 Appendix Table 4–25; Survey of Earned Doctorates, Table 12).

increased (Agrawal and Goldfarb 2008; Wagner, Park, and Leydesdorff 2015). In fact, multi-university teams are the fastest growing authorship structure (Jones, Wuchty, and Uzzi 2008). In addition, distant collaboration is one teamwork feature where economics and the social sciences appear ahead of the hard sciences: Collaborators in the hard sciences and engineering are more likely to remain co-located, perhaps due to a greater reliance on capital equipment (Jones et al. 2008).

A rising frequency of teams also follows naturally from declining collaboration costs. Declining communications costs provide less direct reasoning for why one collaborates in the first place or why the impact of teams is rising, especially in dominating the upper tail of scientific and social scientific work. This suggests that one should understand the evolution of teamwork as a mix between the costs of collaboration and its innovative benefits, where easing collaboration through technology makes it easier for individuals to seize these advantages. Relatedly, much of the early Internet itself was developed through the National Science Foundation to facilitate research, so that the reduction in communication costs has followed in part from the fruits of collaboration that researchers perceived (for example, Greenstein 2010).

A separate dimension of "communication" is that teamwork may provide promotional advantages. That is, the diffusion of an idea may be increasing in the number of authors who presumably extend the promotional opportunities for a given paper. This "marketing" advantage may be an additional, longstanding benefit of working in a team. However, it seems less useful for explaining the rising impact advantage of teams of a given size or the sharply increasing advantage of teams among the very highest impact work.

Summary

Integrating across these perspectives, several interrelated themes emerge. In part, teamwork is an exercise in accessing horizontally differentiated information. The aggregation of expertise can in turn improve creative search and implementation amidst the large and expanding landscape of existing knowledge. These perspectives are consistent with various empirical evidence, from direct surveys of why people form research teams to psychology experiments around team function to big-data descriptive findings. These perspectives also correspond to straightforward conceptual reasoning, linking the cumulativeness of ideas to inevitable individual narrowness. At the same time, teamwork is increasingly advantaged by how improved communication technologies reduce collaboration costs. This advance has allowed teamwork not only to increase locally but also to draw together researchers working at large geographical distances. While these considerations are not a comprehensive picture of the underlying forces in the rise of teams, they engage the remarkable generality in the rise of teams across economic subfields and virtually all scientific, social scientific, and engineering research disciplines. These dimensions also point to a specific series of challenges facing the economics field, to which we turn next.

Team Costs: Challenges for Individuals and Challenges for the Field

Given the increasing impact advantage of research teams, one might conclude that the shift toward teamwork is a fruitful development for the field. However, the rise of teams can also create costs, especially for the development of scientific careers. Here we consider key dimensions of these costs as well as interventions that might lessen the challenges.

Teamwork and the Academic Reward System

The progress of scholarly careers depends on community perception. And the key input to community perception is the scientific work that scholars produce. Merton (1957) describes the "thin" property right of scientific ideas, where authors place their work in the public domain and are rewarded especially strongly according to community-level assessment, both about the quality and originality of the idea and the role of the specific scientist in its creation. As Merton writes, "In short, property rights in science become whittled down to just this one: the recognition by others of the scientist's distinctive part in having brought the result into being." In the academic reward system, high-stakes decisions on tenure, grant funding, journal placement, presentation opportunities, the awarding of prizes, the attachment of one's name to an idea (eponymy), and status itself all rely on formal and informal evaluation by the community. During the era of solo research, crediting individuals for their work was a relatively straightforward exercise. The rise of teams clouds community inference, which raises difficult and potentially pernicious issues.

Consider the tenure decision. Lifetime contracts are typically awarded, or not, based largely on publications in a handful of years after the PhD, with the duration specified by the institution's "tenure clock." What publication information is available to make this decision? Figure 4 charts the evolution of early-career publishing, drawing from all economics articles indexed in the Microsoft Academic Graph and focusing on individuals who publish multiple papers. We define cohorts of economists based on their first publication year and count their publications through the ensuing seven years. Further, we consider publication counts for individuals who achieve different percentiles of lifetime citations. Again, the percentile thresholds for lifetime citations are determined within cohorts: for example, the "upper 10 percent" for the year 1970 considers, among all the people who first published in 1970, the 10 percent of individuals with the highest citation counts through 2018.

Figure 4, panel A shows that solo-authored work in the early career has become extremely rare. By 2010, the typical early publication record shows just one solo publication. This (new) regularity appears across economists of different impact profiles and regardless of how prolific they are in general. The decline of solo-authored work comes despite an increase in overall publication counts shown in Figure 4, panel B. The early-career publication record of economists can thus provide a number of signals, but these signals are increasingly intertwined. When the single solo-authored work is a job-market paper, the individual's additional

Figure 4
Authorship in the Early Career

Panel A. Count of solo-authored papers in first eight years of career
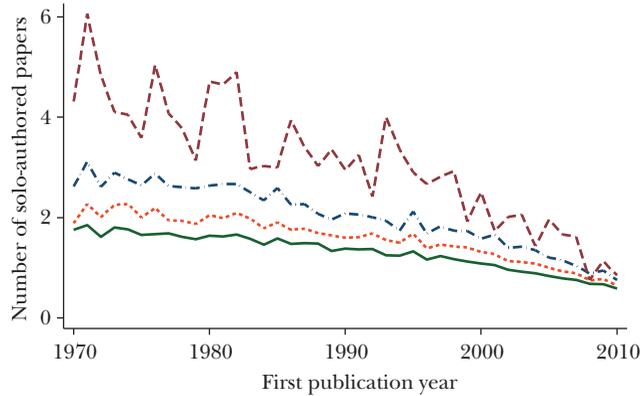
Panel B. Count of all papers in first eight years of career

--- Upper 1%   – – Upper 10%   ······· Upper 25%   —— Upper 50%
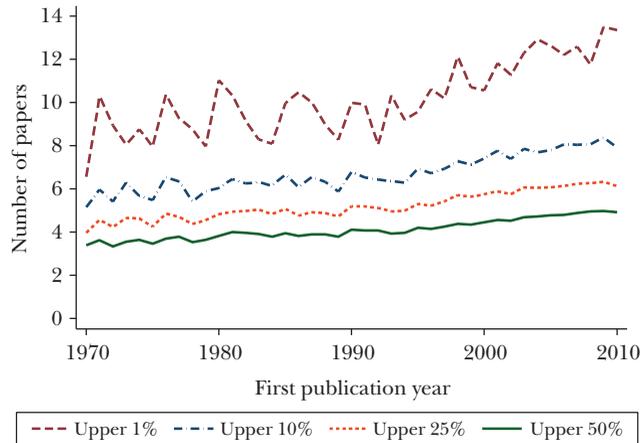
Note: Cohorts are defined by the individual's first publication year. The "Upper N%" subgroups are defined according to each person's career total citations, with comparisons made within the cohort. See text for further detail. For details, see the online Appendix available with this article at the *JEP* website.

pre-tenure publication signals are entirely mixed with their coauthors. Tenure decisions must now rely on credit assignment within teams.

Next, consider scholarly prizes. Individual prizes are highly sought after in many fields, including economics, where the Nobel prize and the Clark medal are extremely prominent.[5] However, the more work that is done in teams, and the

---

[5] The Nobel prize can be awarded to up to three people in a given year and can, in principle, be given to a research team. While this happens in the sciences, where the award is typically given for a particular breakthrough, in economics the Nobel is typically given to selected individuals for their broader bodies

more impactful this work becomes, the less obvious the decisions become on whom specifically to reward. By contrast, "best paper" prizes, such as the Frisch medal given by the econometric society, do not require a parsing of credit. Prizes for "best paper" may then be increasingly appropriate reward mechanisms as the nature of economics and science more broadly continues to shift.

Overall, community evaluators are left in the high-stakes but increasingly untethered role of assigning credit. To some extent, one might make additional, individually informed inferences to help settle the matter. For example, perhaps a series of papers look like one author's agenda, as opposed to another's agenda. Or perhaps the paper mixes techniques (say, theory and empirical work), where each author can be assigned to particular techniques, and these techniques can be evaluated separately within the paper. But often such inferences are difficult. With nearly all work being team-authored, it is increasingly difficult to identify the contribution of each individual author. It is even more challenging in the early career before tenure—before individuals have known agendas or technical strengths.

Credit and Bias

More pernicious problems may also fester within the informational voids. In assigning credit across team members, community members may consciously or subconsciously weight their views based on gender, race, or other group characteristics, including the institutional affiliations of the authors. When individual signals are weakened, the role of people's priors or group preferences are comparatively strengthened. These issues are not just theoretical in economics: Sarsons et al. (2021) consider the high-stakes outcome of tenure promotion comparing female and male economists. They find that women are penalized for coauthored work in tenure decisions, while their male counterparts are not. The rise of teams may thus, inadvertently, worsen well-known problems of underrepresentation and discrimination within the economics field (Bayer and Rouse 2016; Allgood, Badgett, and Bayer 2019).

The assignment of credit at one point in time also has future career implications. If one person receives too little credit for a given work, someone else receives too much. In addition, the more the community believes one person deserves credit, the more advantages that person may receive in the future. This idea, coined the "Matthew Effect" in science (Merton 1968), emphasizes a success-begets-success or rich-get-richer and poor-get-poorer dynamic within research domains.[6]

Bias in credit can translate into bias in funding resources and access to top journals, prizes, and the academy's other attendant rewards. Natural experiments in a variety of settings outside economics show that eminence, including both personal

---

of work. Similarly, the Clark medal can, in principle, be awarded jointly to two people, where their work is intertwined, but to date this has not been done.

[6]The Matthew Effect is named in accordance with the Biblical passage, "For to everyone who has will more be given, and he will have abundance; but from him who has not, even what he has will be taken away" (Matthew 25:29, Revised Standard Version).

eminence and institutional eminence, draws favorable attention by reviewers and in the diffusion of one's ideas (for example, Simcoe and Waguespack 2011; Azoulay, Stuart, and Wang 2014; Hill and Stein 2020) and that early funding advantages in the sciences lead to later funding advantages (Bol, de Vann, and de Rijt 2018). Moreover, teams do not share credit equally. For example, in the context of scientific retractions across the sciences, Jin et al. (2019) find that more eminent coauthors see little decline in future citations to their work after a retraction, but the junior coauthors do; in fact, eminent authors are protected more strongly when junior coauthors are involved in the project.

Teamwork thus has the potential to worsen bias, potentially in ways with long-run effects. These issues are not just consequential for individuals, but also for the progress of ideas, especially to the extent that resources are misallocated, and talented individuals receive fewer opportunities, depart the field, or, anticipating bias, do not enter in the first place (Phelps 1972; Arrow 1973; Buckles 2019; Hsieh et al. 2019).

The Access Problem

The credit problem arises once a work has been produced. Related challenges also come before work is produced as part of a team. In hitching one's production (and hence career progress) to other individuals, scholars may be highly concerned about the quality of specific teammates and the effort they will provide (Holmstrom 1982; Ahmadpoor and Jones 2019). Especially with complementary tasks (Jones 2009; Lee, Walsh, and Wang 2015), one has to be careful in the choice of coauthors. Mistakes become very costly.

Various evidence indicates that individuals do take care in forming teams. Rather than random selection of team members, scholars in all fields engage in positive assortative matching (Ahmadpoor and Jones 2019). They also rely on close personal contact. For example, while collaboration at a geographic distance is rising, the large majority of such collaboration occurs between individuals who were previously co-located in the same institution (Freeman, Ganguli, and Murciano-Goroff 2015).

In this context, the same forces that drive "credit bias" for work that has already been published can appear again as "access bias" in team assembly. Confronted with potential teammates, prior beliefs about the PhD institution, gender, race, or some other characteristic believed to inform the distribution of quality may strongly influence the demand for that person as a coauthor. These external characteristics will likely play stronger roles when that person is new to their career or when they are not personally known by the other team members.

An increasing division of labor may worsen these issues. When individuals bring differentiated skills or expertise, it is hard for one specialist to evaluate the quality of an individual in a different specialty. In this setting, people may increasingly rely on external signals (like the research institution) to locate collaborators. In fact, we see trends in this direction across the sciences and social sciences amidst the growth in long-distance collaborations. Long-distance collaborations are dominated by elite

institutions, where authors increasingly collaborate with each other and increasingly do not collaborate with authors at lower-ranked institutions (Jones, Wuchty, and Uzzi 2008). The American Economic Association's recent survey found widespread concerns about elitism in the field and summarized key findings as follows: "There is a strong sense that the AEA, the NBER, and the top journals—and de facto the profession—are controlled by economists from the top institutions. Many comments (about 60) reference that there is a network, and you are either in or out, and if you are out, your chances of success in the profession are greatly diminished" (Allgood, Badgett, and Bayer 2019).[7]

Ultimately, teamwork has proven to be a highly stratified activity. While positive assortative matching can be efficient in some static models, obstacles to career access and progression built on discrimination suggest a much more negative view. Discriminatory losses from misallocating talent can be vast (Hsieh et al. 2019). Separately, from efficiency issues, the inequities of discrimination are first-order concerns. Overall, limiting the prospects for groups of talented people, or losing them entirely, can be especially consequential for both individuals themselves and for the collective progress of economics research.

Institutions and Opportunities

The rise of teams is a powerful shift in economics. The choice to work in teams appears natural given its rising impact advantage, and, as discussed above, there are a number of reasons to see the collaborative form as an increasingly important way to drive successful economics research. At the same time, teamwork strains the reward system of science. The decline of transparent signals about individual-level output leaves the community with an increasingly murky challenge in deciding whom to promote and reward. It may also further stratify the field, and the informational voids can interact especially badly with discriminatory preferences or beliefs.

Given this tension, policy interventions must proceed with care. If returning to solo work in economics would reduce research productivity and slow the progress of ideas, solutions will ideally work within the rise of teams. The institutions of economics—universities, journals, funders, and professional associations—all have some power to experiment with interventions that may ameliorate the problems that teams impose.

First, economics journals can consider publication rules that help clarify individual contributions. Outside economics, author order is often used to signal relative contributions. For example, the first and/or last position in the author list, depending on the field, can be used to communicate elevated roles within the

---

[7] Initial discrimination in team assembly can also naturally have dynamic consequences, continuing to limit individuals in the set of coauthors that are willing to match. Team production makes it increasingly difficult to develop a signal of one's talent and change community perception. To the extent that team output follows the harmonic average of the individual productivities (Jones and Ahmadpoor 2019), the fact that an individual may be more talented than their teammates becomes muted in the joint output. This further reduces the community's ability to detect the individual's talent and can leave the community to its prior and potential biases.

team. Further, certain leading science journals now require authors to assert (in a front-page footnote) each person's specific contributions to the work, which might include statistical modeling, formal theory, conceptual development, data acquisition, and writing. If coauthors overlap on many tasks, which may often be the case in economics, stating this fact can still be informative. Such policies may be useful for economics to consider. To put it more strongly: the reasoning for why economics should do things differently from the hard sciences, which have actively engaged these issues, is not obvious. One caveat may be that, within hierarchical teams where PhD students or post-docs work for their advisor or employer, or where junior faculty work with senior colleagues, the junior authors may feel pressure to be overly generous in crediting the senior author. Studies of the effects of these author order systems will be helpful.

Second, funders should be mindful about maintaining access, especially for junior scholars. The more that economists work in teams, the longer it takes to establish an independent reputation and the harder it may be to convince a funder (and the reviewers of grant proposals) that the investment is worthwhile. As a cautionary tale, consider biomedicine, where team sizes have grown large and it is difficult to develop an independent record. The average age at first grant from the National Institutes of Health rose to 43 in 2016, up from age 36 in 1980 (Mann 2017). Former NIH director Elias Zerhouni saw the trend against younger researchers as a major crisis, caused by bias within peer review, that would choke off the pipeline of talent (Kaiser 2008). To the extent that grants become increasingly important to fund team-based economics research, similar concerns may apply. This concern may be especially germane for resource-intensive "laboratory model" style research, which, as discussed above, appears to be increasingly prominent in economics research. Targeting grants at younger scholars may encourage career development and access to these models: indeed, the NIH has responded with quotas to support younger principal investigators. Even if such grants are less-informed bets, they may be important dynamically for the health of the field.

Third, proactive steps can strengthen individual-level assessment opportunities. Seminar and conference presentations can give greater visibility to less-established authors on a team. To the extent that coauthors have all made large contributions to a paper, inviting the less established authors would presumably provide the same valuable research interactions. Similarly, even short visiting opportunities can extend networks for less established scholars. For conference organizers, including discussants can become a universal norm, motivated not just by its benefits for addressing the ideas, but as an opportunity for individuals—and especially less established individuals—to showcase themselves. As team-authored work takes over, such individual-level opportunities become especially useful.

Finally, amidst the rise of teams, economics should work toward objective methods for assessing individual performance of those working in teams. For example, some review panels "divide by N" when crediting team-authored works to individuals, while others wholly credit each author for the work, and still other institutions do not use explicit rules. Reviewers may engage a kind of "fixed effects"

reasoning, attempting to assess an individual by looking at what happens when that individual joins projects, and adjudicate credit based on how the coauthors perform in their other work. In all cases, the concern is that when procedures are typically done non-transparently and in an ad hoc manner, reviewers are more likely to find themselves drawing on personal biases. Ahmadpoor and Jones (2019) develop an explicit method for calculating individual fixed effects, using the citations each paper receives to estimate individual fixed effects, based on the entire collaborative network of the field and looking at all fields in the sciences and social sciences, including economics. Using publication records, the algorithm produces a measure of the citations each author would be expected to produce should that author write alone. In out-of-sample tests, the method is substantially more accurate than other existing approaches for rating individuals. Advancing this type of method may better ground community assessments and limit credit bias (if not access bias), given the rise of teams.[8]

Beyond explicit assessment methodologies to confront the rise of teams, training and procedures to avoid implicit bias may be very important, as one should be mindful of the limits of one's capacity to apportion credit fairly when making judgments about individuals. When serving on review committees, the committee can be explicit about how the group is apportioning credit amidst team-authored work. It should be standard for reviewers to ask each other to explain the basis for their judgments. This kind of "peer review within peer review" can help the reviewers perceive and limit any bias.

Finally, there is the training aspect for team members themselves. Working in teams in economics is currently a learning-by-doing affair. But teamwork engages collaborative and communications skills and technological platforms that are distinct from the methodological tools and domain knowledge one is formally taught in a PhD program. Hierarchical teamwork and laboratory approaches further engage management skills and, often, skills at acquiring funding. PhD training to advance these skills would provide increasingly valuable inputs to economists' careers, and various team-skill schematics and training efforts in the sciences can provide models (see National Research Council 2015).

Conclusion

The traditional image of economic research involves someone working alone: one pictures Adam Smith writing *An Inquiry into the Nature and Causes of the Wealth of Nations* in the 18th century, Karl Marx sitting in the reading rooms of the British

---

[8]By focusing purely on the publication record, the method avoids direct consideration of characteristics that may be discriminatory. However, to the extent that citations themselves reflect community biases, this method may still indirectly imprint bias onto the individual measure. Assessing and correcting for any such indirect bias (which can be relevant in machine learning and other statistical prediction approaches) is an important additional step.

Museum in the 19th century, or Joan Robinson writing *The Economics of Imperfect Competition* in the 1930s, all as solitary endeavors. The evidence presented in this paper suggests that the traditional view was once largely accurate but has become increasingly outdated. Economics, like many other areas of academic research, has been moving to a team-based approach. A modern view of academic research in economics would focus on how research teams are formed, how the teams evolve over time, how they are funded, how they function, how they are managed, and how the economics community ascribes credit to the individuals who participate in them. This shift in perspective from individual to team-based research should have implications for what it means to train an economist, what the career path for an economics researcher looks like, and how a research trajectory will be rewarded.

Maintaining the benefits of teamwork while managing the challenges is not simple. A broad view is that the economics profession can look to the sciences, given their longer-standing team orientation, for policy ideas. Several policy innovations that may manage the challenges have been articulated here. Improvements in institutional design will also greatly benefit from careful experimentation and further study. It is certainly time to address these issues, especially given the issues of bias in economics, and as the prevalence and impact of teamwork in economic research continues to rise.

References

Adams, James D., Grant C. Black, J. Roger Clemmons, and Paula E. Stephan. 2005. "Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities, 1981–1999." *Research Policy* 34 (3): 259–85.

Agrawal, Ajay, and Avi Goldfarb. 2008. "Restructuring Research: Communication Costs and the Democratization of University Innovation." *The American Economic Review* 98 (4): 1578–90.

Agrawal, Ajay, Avi Goldfarb, and Florenta Teodoridis. 2016. "Understanding the Changing Structure of Scientific Inquiry." *American Economic Journal: Applied Economics* 8 (1): 100–128.

Ahmadpoor, Mohammad, and Benjamin F. Jones. 2019. Decoding Team and Individual Impact in Science and Invention. *PNAS* 116 (28): 13885–90.

Allgood, Sam, and Lee Badgett, Amanda Bayer, Marianne Bertrand, Sandra E. Black, Nick Bloom, and Lisa D. Cook. 2019. *AEA Professional Climate Survey: Final Report.* Nashville, TN: AEA.

Arrow, Kenneth. 1973. "The Theory of Discrimination." In *Discrimination in Labor Markets*, edited by Orley Ashenfelter and Albert Rees, 3–33. Princeton: Princeton University Press.

Azoulay, Pierre, Toby Stuart, and Yanbo Wang. 2014. Matthew: Effect or Fable? *Management Science* 60 (1): 92–109.

Bayer, Amanda, and Cecilia Elena Rouse. 2016. "Diversity in the Economics Profession: A New Attack on an Old Problem." *Journal of Economic Perspectives* 30 (4): 221–42.

Becker, Howard S. *Art Worlds.* 1982. Berkely: University of California Press.

Bol, Thijs, Mathijs de Vaan, and Arnout van de Rijt. 2018. "The Matthew Effect in Science Funding." *PNAS* 115 (19): 4887–90.

Buckles, Kasey. 2019. "Fixing the Leaky Pipeline: Strategies for Making Economics Work for Women at Every Stage." *Journal of Economic Perspectives* 33 (1): 43–60.

Cole, John Y. 1996. "Publishing at the Library of Congress: A Brief History." *Publishing Research Quarterly* 12 (2): 38–49.

Currie, Janet, Henrik Kleven, and Esmée Zwiers. 2020. "Technology and Big Data Are Changing Economics: Mining Text to Track Methods." *AEA Papers and Proceedings* 110: 42–48.

Einstein, Albert. 1941. *The World as I See It.* London: John Lane the Bodley Head.

Fleming, Lee. 2001. "Recombinant Uncertainty in Technological Search." *Management Science* 47 (1): 117–32.

Freeman, Richard B., Ina Ganguli, and Raviv Murciano-Goroff. 2015. "Why and Wherefore of Increased Scientific Collaboration." In *The Changing Frontier: Rethinking Science and Innovation Policy*, edited by Adam B. Jaffe and Benjamin F. Jones, 17–48. Chicago: University of Chicago Press.

Funk, Russell J., and Jason Owen-Smith. 2016. "A Dynamic Network Measure of Technological Change." *Management Science* 63 (3): 791–817.

Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *The Quarterly Journal of Economics* 114 (3): 739–67.

Greenstein, Shane. 2010. "The Emergence of the Internet: Collective Invention and Wild Ducks. Industrial and Corporate Change." 19 (5): 1521–62.

Hill, Ryan, and Carolyn Stein. 2020. "Scooped! Estimating Rewards for Priority in Science." https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjr_Ye84LXvAhXhdM0KHWEbDv4QFjAAegQIAxAD&url=http%3A%2F%2Feconomics.mit.edu%2Ffiles%2F20197&usg=AOvVaw084tisiS8u9c0yJ5rG-yck.

Holmstrom, Bengt. 1982. "Moral Hazard in Teams." *Bell Journal of Economics* 13 (2): 324–40.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. 2019. "The Allocation of Talent and U.S. Economic Growth." *Econometrica* 87 (5): 1439–74.

Inwood, Stephen. 2003. "The Man Who Knew too Much: The Strange and Inventive Life of Robert Hooke, 1635–1703." London: Pan Macmillan.

Jin, Ginger Zhe, Benjamin Jones, Susan Feng Lu, and Brian Uzzi. 2019. "The Reverse Matthew Effect: The Consequences of Retraction in Scientific Teams." *Review of Economics and Statistics* 101 (3): 492–506.

Johnson, Rob, Anthony Watkinson, and Michael Mabe. 2018. *The STM Report: An Overview of Scientific and Scholarly Publishing*. Tha Hague: International Association of Scientific, Technical and Medical Publishers.

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.

Jones, Benjamin F. 2010. "Age and Great Invention." *The Review of Economics and Statistics* 92 (1): 1–14.

Jones, Benjamin F. 2014. "The Knowledge Trap: Human Capital and Development Reconsidered." Northwestern University Working Paper.

Jones, Benjamin F., Stefan Wuchty, and Brian Uzzi. 2008. "Multi-University Research Teams: Shifting Impact, Geography and Stratification in Science." 322 (5905): 1259–62.

Kaiser, Jocelyn. 2008. "Zerhouni's Parting Message: Make Room for Young Scientists." *Science* 322 (5903): 834–35.

Kaldor, Nicholas. 1961. "Capital Accumulation and Economic Growth." In *The Theory of Capital*, edited by F.A. Lutz and D.C. Hague, 177–222. New York: St. Martins Press.

Kim, Daniel, Daniel Burkhardt Cerigo, Hawoong Jeong, and Hyejin Youn. 2016. "Technological Novelty Profile and Invention's Future Impact." *EPJ Data Science* 5 (8).

Lee, You-Na, John P. Walsh, and Jian Wang. 2015. "Creativity in Scientific Teams: Unpacking Novelty and Impact." *Research Policy* 44 (3): 684–97.

Mann, Sarah. 2017. "NIH, Research Community Target Next Generation of Scientists," November 13. https://www.aamc.org/news-insights/nih-research-community-target-next-generation-of-scientists.

Merton, Robert K. 1957. "Priorities in Scientific Discovery: A Chapter in the Sociology of Science." *American Sociological Review* 22 (6): 635–59.

Merton, Robert K. 1968. "The Matthew Effect in Science." *Science* 159 (3810): 56–63.

Mesmer-Magnus, Jessica R., and Leslie A. DeChurch. 2009. "Information Sharing and Team Performance: A Meta-Analysis." *Journal of Applied Psychology* 94 (2): 535–46.

Morrison, Samuel Eliot. 1936. *Three Centuries of Harvard: 1636–1936.* Cambridge, MA: Harvard University Press.

Mukherjee, Satyam, Daniel M. Romero, Ben Jones, and Brian Uzzi. 2017. "The Nearly Universal Link between the Age of Past Knowledge and Tomorrow's Breakthroughs in Science and Technology: The Hotspot." *Science Advances* 3 (4).

National Research Council of the National Academies. 2015. *Enhancing the Effectiveness of Team Science.* Washington, DC: National Academies Press.

Oettl, Alexander. 2012. "Reconceptualizing Stars: Scientist Helpfulness and Performance." *Management Science* 58 (6): 1122–40.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62 (4): 659–61.

Rothenberg, Albert. 2015. *Flights from Wonder: An Investigation of Scientific Creativity.* Oxford: Oxford University Press.

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram. 2021. "Gender Differences in Recognition for Group Work." *Journal of Political Economy* 129 (1): 101–47.

Schilling, Melissa. A. 2005. "A 'Small World' Network Model of Cognitive Insight." *Creativity Research Journal* 17: 131–54.

Schumpeter, Joseph A. 1939. *Business Cycles.* New York: McGraw-Hill Book Company.

Schweitzer, Sascha, and Jan Brendel. 2019. "A Burden of Knowledge Creation in Academic Research: Evidence from Publication Data." https://ssrn.com/abstract=2895685.

Simcoe, Timothy S., and Dave M. Waguespack. 2011. "Status, Quality, and Attention: What's in a (Missing) Name?" *Management Science* 57 (2): 274–90.

Sinha, Arnab, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June (Paul) Hsu, and Kuansan Wang. 2015. "An Overview of Microsoft Academic Service (MAS) and Applications." http://dx.doi.org/10.1145/2740908.2742839.

Usher, Abbott Payson. 1954. *A History of Mechanical Invention.* Cambridge, MA: Harvard University Press.

Uzzi, Brian, and Jarrett Spiro. 2005. "Collaboration and Creativity: The Small World Problem." *American Journal of Sociology* 111 (2): 447–504.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 25: 468–72.

Wagner, Caroline S., Han Woo Park, and Loet Leydesdorff. 2015. "The Continuing Growth of Global Cooperation Networks in Research: A Conundrum for National Governments." *PloS one* 10 (7).

Weitzman, Martin L. 1998. "Recombinant Growth." *Quarterly Journal of Economics* 113 (2): 331–60.

Wu, Lingfei, Dashun Wang, and James A. Evans. 2019. "Large Teams Develop and Small Teams Disrupt Science and Technology." *Nature* 566: 378–82.

Wuchty, Stefan, Benjamin F. Jones, and Brian Uzzi. 2007. "The Increasing Dominance of Teams in the Production of Knowledge." *Science* 316 (5827): 1036–39.

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of under-graduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by e-mail at taylort@macalester.edu, or c/o *Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., St. Paul, MN 55105.

Smorgasbord

Raj Chetty delivered a keynote address, "Improving Equality of Opportunity: New Insights from Big Data," to the annual meetings of the Western Economic Association International (*Contemporary Economic Policy*, January 2021, 39 (1): 7–41, https://onlinelibrary.wiley.com/doi/10.1111/coep.12478, subscription needed). "[T]he geographic scale on which we should think about neighborhoods as they matter for economic opportunity and upward mobility is incredibly narrow, like a half mile radius around your house. We find this not just for poverty rates, but many other characteristics. If you look at differences in characteristics outside that half mile radius, they have essentially no predictive power at all. I think that's extremely useful from a policy perspective. We started this talk with the American dream. We now see that its origins, its roots, seem to actually be extremely hyperlocal."

Robert J. Shiller delivered the Godley-Tobin Lecture at the Eastern Economic Association meetings, on the topic, "Animal Spirits and Viral Popular Narratives" (*Review of Keynesian Economics*, January 2021, 9 (1): 1–10, https://www.elgaronline.com/view/journals/roke/9-1/roke.2021.01.01.xml). "Mathematical epidemiology has been studying disease phenomena for over a century, and its frameworks can provide an inspiration for improvement in our understanding of economic dynamics. People's states of mind change through time, because ideas can be contagious, so that they spread from person to person just as diseases do. . . . We humans live our lives in a sea of epidemics all at different stages, including epidemics of diseases and epidemics of narratives, some of them growing at the moment, some peaking at the moment, others declining. New mutations of both the diseases and the narratives are constantly appearing and altering behavior. It is no wonder that changes in business conditions are so often surprising, for there is no one who is carefully monitoring the epidemic curves of all these drivers of the economy. Since the advent of the internet age, the contagion rate of many narratives has increased, with the dominance of social media and with online news and chats. But the basic nature of epidemics has not changed. Even pure person-to-person word-of-mouth spread of epidemics was fast enough to spread important ideas, just as person-to person contagion was fast enough to spread diseases into wide swaths of population millennia ago."

Lars Vilhuber discussed "Reproducibility and Replicability in Economics" (*Harvard Data Science Review*, Fall 2020, https://hdsr.mitpress.mit.edu/pub/fgpmpj1l/release/3). "In 1960, 76% of empirical *AER* [*American Economic Review*] articles used public-use data. By 2010, 60% used administrative data, presumably none of which is public use . . ." "Still, after 30 years, the results of reproducibility studies consistently show problems with about a third of reproduction attempts, and the increasing share of restricted-access data in economic research requires new tools, procedures, and methods to enable greater visibility into the reproducibility of such studies."

Christopher J. Coyne and Peter J. Boettke offer an introduction to *The Essential Austrian Economics* (Fraser Institute 2020, https://www.essentialscholars.org/austrian-economics). "It is only by allowing decentralized people to participate in an ongoing process of discovery that the knowledge necessary to make rational economic decisions emerges. These numerous discoveries lead to the emergence of knowledge regarding not only what goods and services are desired by consumers, but also the most effective techniques to produce these outputs in a cost-minimizing manner. The problems inherent with market socialism, according to Hayek, were not a matter of placing smarter people in charge or in developing new computational techniques to gather more information. Instead, the issue was that the economic knowledge necessary for coordination is dispersed, tacit, and emergent. This means that the knowledge used by people to coordinate their economic affairs cannot exist outside the context within which they are embedded. The market socialism model left no space for the very activity that generated the knowledge that was necessary for planners to accomplish their stated ends of advanced material

production. . . . The emphasis on the division of knowledge and the market process as a means of discovering and using this knowledge is the crux of the Austrian criticism of both comprehensive and piecemeal government intervention into a freely operating market."

Russell Pittman discusses "On the Economics of Restructuring World Railways, with a Focus on Russia" (January 2021, US Department of Justice, Economic Analysis Group Working Paper 21–1, https://www.justice.gov/atr/page/file/1358981). A version of this paper is also published in *Man and the Economy* (December 2020, 7 (2), https://www.degruyter.com/document/doi/10.1515/me-2020-0014/html, subscription required). "The European Commission has been very strong on pushing complete vertical separation: competition above the rail among independent train operating companies. . . . On the other hand, in the Americas, North and South America and Central America, we have almost exclusively horizontal separation. Competition among vertically integrated railway companies that own their track in the U.S. and Canada, or have long-termfranchise control of their track in Mexico and Brazil, and can for the most part insist that only their trains run on their tracks. For the most part, they have the complete right to deny other trains access. . . . However, as we have found out very well, each of these two basic solutions has an Achilles heel. In the EU, . . . the most serious weakness has been the unreliability of public funding of infrastructure. . . . The result has been bottlenecks, lack of expansion where it's needed, slow and unreliable service in many countries in the EU. It's a very big problem. . . . In the Americas, . . . [T]he weakness there has been that every railway has some degree of regional monopoly power. . . . On the other hand, attracting private investment is a strength of the horizontal separation model. . . . The U.S. freight railways . . . are profitable; they earn money; and they spend money on their infrastructure. It's been a real success story."

Eduardo Engel, Ronald Fischer, and Alexander Galetovic discuss "Public–Private Partnerships: Some Lessons After 30 Years" (*Regulation*, Fall 2020, pp. 30–35, https://www.cato.org/sites/cato.org/files/2020-09/regulation-v43n3-2.pdf). The subheading reads: "The savings policymakers usually claim for these projects are illusory, but well-designed contracts can deliver public benefits." "[I]nvestment in PPPs over the last 30 years has been substantial, adding €203 billion of infrastructure spending in Europe and $535 billion of spending in developing countries. Most investments are in roads, seaports, and airports, but in some countries investment via PPPs has been significant in other types of infrastructure, such as hospitals and schools. In comparison, PPP investments in the United States have been small." This article complements the discussion in this issue of *JEP* about the Odebrecht case by these three authors and Nicolas Campos.

Jane Ihrig and Scott Wolla recommend, "Let's close the gap: Revising Teaching Materials to Reflect How the Federal Reserve Implements Monetary Policy" (October 2020, Federal Reserve Finance and Economics Discussion Series 2020–092, https://www.federalreserve.gov/econres/feds/files/2020092pap.pdf). "Over the past decade or so, the Fed has purposefully changed the way it implements monetary policy. Unfortunately, many teaching resources have not been updated. Before the

financial crisis of 2007–2008, the Fed implemented policy with *limited* reserves in the banking system and relied on the daily use of open market operations as its key tool. Today and over the longer run, the Fed has stated that it plans to implement policy with *ample* reserves and rely on its administered interest rates. These changes, along with a few others, seem subtle, but the current framework is very different from the previous one. And, these changes are not well reflected in teaching resources."

Symposia and Collections

Melissa S. Kearney and Amy Ganz have edited a nine-chapter book *Securing Our Economic Future*, with sections on Economics of the American Middle Class, the Geographic Disparities in Economic Opportunity, and the Geopolitics of the Climate and Energy Challenge and the US Policy Response (Aspen Institute Economic Strategy Group, December 2020, https://www.economicstrategygroup. org/publication/securing_our_economic_future). As one example, David Autor discusses "The Faltering Escalator of Urban Opportunity." "In the initial decades following WWII, U.S. cities offered a distinctive skills and earnings escalator to less-educated workers. A likely reason why is that, in these decades, adults without college degrees performed higher-skilled, more specialized jobs in cities than their non-urban counterparts. Laboring in urban factories and offices, they staffed middle-skill, middle-pay production, clerical, and administrative roles, where they worked in close collaboration with highly educated professionals (e.g., engineers, executives, attorneys, actuaries, etc.). These collaborative working relationships often demanded specific skills and shared expertise, and likely contributed to the higher wages (and higher productivity) of urban non-college workers. . . . In the decades since 1980, however, this distinctive feature of urban labor markets has diminished. As rising automation and international trade have encroached on employment in urban production, administrative support, and clerical work, the noncollege urban occupational skill gradient has diminished and ultimately disappeared. While urban residents are on average substantially more educated—and their jobs vastly more skill-intensive—than four decades ago, non-college workers in U.S. cities perform substantially less specialized and more skill-intensive work than they did decades earlier. Polarization thus reflects an unwinding of the distinctive structure of work for non-college adults in dense cities and metro areas relative to suburban and rural areas. And as this distinctive occupational structure has receded, so has the formerly robust urban wage premium paid to non-college workers."

The UK Wealth and Policy Commission has produces a final report, *A Wealth Tax for the UK* by Arun Advani, Emma Chamberlain, and Andy Summers, recommending a one-time wealth tax. The commission has also published nearly 40 background papers on aspects of wealth taxes and their application in difference countries (December 2020, https://www.wealthandpolicy.com/). In Working Paper 106, for example, Sarah Perret asks "Why did other wealth taxes fail and is this time different?" "In 1990, there were twelve OECD countries, all in Europe, that

levied individual net wealth taxes. However, most of them repealed their wealth taxes in the 1990s and 2000s, including Austria (in 1994), Denmark and Germany (in 1997), the Netherlands (in 2001), Finland, Iceland, and Luxembourg (in 2006) and Sweden (in 2007). Iceland, which had abolished its wealth tax in 2006, reintroduced it as a temporary 'emergency' measure between 2010 and 2014. Spain, which had introduced a 100% wealth tax reduction in 2008, reinstated the wealth tax in 2011. The reinstatement of the wealth tax was initially planned to be temporary but has been maintained since. France was the last country to repeal its wealth tax in 2018, replacing it with a tax on high-value immovable property. In 2020, Norway, Spain and Switzerland were the only OECD countries that still levied individual net wealth taxes."

The *University of Pennsylvania Law Review* (June 2020, https://www.pennlawreview.com/print/volume-168/issue-7/) has published a nine-paper symposium on antitrust law, with many of the contributions by economists. As one example, Nancy L. Rose and Jonathan Sallet discuss "The Dichotomous Treatment of Efficiencies in Horizontal Mergers: Too Much? Too Little? Getting it Right." "Here is a stylized example of the role that efficiencies might play in an antitrust review. Imagine two paper manufacturers, each with a single factory that produces several kinds of paper, and suppose their marginal costs decline with longer production runs of a single type of paper. They wish to merge, which by definition eliminates a competitor. They justify the merger on the ground that after they combine their operations, they will increase the specialization in each plant, enabling longer runs and lower marginal costs, and thus incentivizing them to lower prices to their customers and expand output. If the cost reduction were sufficiently large, such efficiencies could offset the merger's otherwise expected tendency to increase prices. . . . As empirically-trained economists focused further on what data revealed about the relationship between mergers and efficiencies, the results cast considerable doubt on post-merger benefits. As discussed at length by Professor Hovenkamp, 'the empirical evidence is not unanimous, however, it strongly suggests that current merger policy tends to underestimate harm, overestimate efficiencies, or some combination of the two.' The business literature is even more skeptical. As management consultant McKinsey & Company reported in 2010: 'Most mergers are doomed from the beginning. Anyone who has researched merger success rates knows that roughly 70 percent of mergers fail.'"

Interviews

The editors of the *Erasmus Journal for Philosophy and Economics*, Akshath Jitendranath and Marina Uzunova have co-authored, "What Egalitarianism Requires: An Interview with John Roemer" (Winter 2020, 13 (2), pp. 127–176, https://ejpe.org/journal/article/view/530/358). As they note: "Roemer's work spans the domains of economics, philosophy, and political science, and, most often, applies the tools of general equilibrium and game theory to problems of political economy and

distributive justice—problems often stemming from the discussions among political philosophers in the second half of the twentieth century." Roemer says: "I believe that all young adults should begin their productive years with the same amount of wealth. This implies that the inheritance of wealth, and *in vivos* transfers to the young, must be sharply constrained. If the educational system has succeeded in eliminating inequality of opportunity, and people make different career choices, then differential wealth will emerge during adult lifetimes, and I believe those differences are consistent with justice, as long as there is sufficient income and wealth taxation to prevent income differences from becoming too extreme—so extreme as to threaten solidarity. . . . Many leftists believe the key to understanding capitalism is to understand the extraction of labor from labor power at the point of production. And indeed, I think Marx sometimes erred in thinking this, as well. My view is that the essence of capitalism is the set of institutions which sanctify and enforce private and unequal ownership of capital—that is, vastly unequal wealth. . . . This is the key locus of power; oppression of workers at the point of production, though perhaps very important in building class consciousness of workers, is relatively small potatoes. Coercion at the point of production was essential in feudalism and slavery, but capitalism has subtler techniques for accumulating wealth. . . . My goal is to focus on building solidaristic societies, and I think that the most important barrier to solidarity is the individualistic ethos of capitalist society where the accumulation of private wealth is the guiding force."

Douglas Clement and Anjali Nair have collaborated on a "Seema Jayachandran interview: On Deforestation, Corruption, and the Roots of Gender Inequality" (Federal Reserve Bank of Minneapolis, February 12, 2021, https://www.minneapolisfed.org/article/2021/seema-jayachandran-interview-on-deforestation-corruption-and-the-roots-of-gender-inequality). "There's a famous U-shaped relationship in the data between economic development and female labor force participation. . . . Historically, in richer countries, you've seen this U-shape where, initially, there are a lot of women working when most jobs are on the family farm. Then as jobs move to factories, women draw out of the labor force. . . . But then there's an uptick where women start to enter the labor market more and not just enter the labor market, but earn more money. There are several reasons why we think that will happen. One is structural transformation, meaning the economy moves away from jobs that require physical strength like in agriculture or mining towards jobs that require using brains. . . . The second reason is improvement in household production. Women do the lion's share of household chores and, as nations develop, they adopt technology that reduces the necessary amount of labor. . . . Some of those technological advances are in infrastructure. Piped water, for instance, where we're relying on the government or others to build that public good infrastructure. And some is within households; once piped water is available, households invest in a washing machine. The third reason is fertility. When countries grow richer, women tend to have fewer kids and have the ability to space their fertility. For example, both the smaller family size and the ability to choose when you have children allows women to finish college before having children."

Discussion Starters

Edward Tenner offers a meditation on "The Importance of Being Unimportant" (*Milken Institute Review*, First Quarter 2021, https://www.milkenreview.org/articles/the-importance-of-being-unimportant). Tenner offers some vivid examples of fortunes made in unimportant inventions, like the tire valve and sewing thread. He discusses Jack and Belle Linsky, the founders of Swingline Corp., who invented the idea that staples could be glued and aligned in a row, rather than being loose individual objects, and the modern stapler to use it. Tenner writes: "Virtually no manufactured object costs less than a staple. Yet this humble device so enriched the Linskys that they were able to compete successfully with the Queen of England in auctions for decorative arts."

Tim Sablik tells how "The COVID-19 pandemic disrupted the supply of many items, including cold hard cash" (*Econ Focus*: Federal Reserve Bank of Richmond, Fourth Quarter 2020, pp. 26–29, https://www.richmondfed.org/publications/research/econ_focus/2020/q4/federal_reserve). "I started getting a few phone calls from members asking, 'Is it just me, or are more quarters walking out the door than before?' says Brian Wallace, president of the Coin Laundry Association. Of the roughly 30,000 self-service laundromats in the United States, Wallace says that a little more than half take only quarters as payment to operate washers and dryers. Before the pandemic, some of these coin-operated businesses would take in more quarters each week than they gave out, meaning that most customers brought their own change to the laundromat rather than exchanging bills for quarters. But as the pandemic intensified, many of those business owners who had been used to ending the week with a surplus of quarters suddenly found they had a deficit. They turned to their local bank to purchase more, but the banks had no change to spare either." Sablik tells how the Fed started rationing coins in June 2020, and how October 2020 was 'get coin moving month': "One aquarium in North Carolina shuttered by the pandemic put its employees to work hauling 100 gallons of coins from one of its water fixtures that had served as a wishing well for visitors since 2006."

Introducing a new look for IMF eLibrary

Same free content with updated features.
eLibrary.IMF.org

The American Economic Association

MIX
Paper from responsible sources
FSC
www.fsc.org
FSC™ C132124

The Journal of
Economic Perspectives

Spring 2021, Volume 35, Number 2

Symposia

European Union
Philip R. Lane, "The Resilience of the Euro"
Keith Head and Thierry Mayer, "The United States of Europe:
A Gravity Model Evaluation of the Four Freedoms"
David Dorn and Josef Zweimüller, "Migration and Labor Market
Integration in Europe"
Florin Bilbiie, Tommaso Monacelli, and Roberto Perotti, "Fiscal Policy
in Europe: Controversies over Rules, Mutual Insurance, and Centralization"

Preventive Medicine
Joseph P. Newhouse, "An Ounce of Prevention"
Amanda E. Kowalski, "Mammograms and Mortality: How Has
the Evidence Evolved?"

Articles
M.V. Lee Badgett, Christopher S. Carpenter, and Dario Sansone,
"LGBTQ Economics"
Nicolás Campos, Eduardo Engel, Ronald D. Fischer, and Alexander Galetovic,
"The Ways of Corruption in Infrastructure: Lessons from the Odebrecht Case"
Benjamin F. Jones, "The Rise of Research Teams: Benefits
and Costs in Economics"

Feature
Timothy Taylor, "Recommendations for Further Reading"

AMERICAN
ECONOMIC
ASSOCIATION