The Journal of

Economic Perspectives

A journal of the
American Economic Association

Winter 2023

The Journal of
Economic Perspectives

A journal of the American Economic Association

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

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

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.
Printed by LSC Communications, Owensville, Missouri 65066, USA.

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Winter 2023, Vol. 37, No. 1. The JEP is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. For details and further information on the AEA go to https://www.aeaweb.org/. Periodicals postage paid at Nashville, TN, and at additional mailing offices.

POSTMASTER: Send address changes to the *Journal of Economic Perspectives*, 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the USA.

The Journal of
Economic Perspectives

Contents

Volume 37 • Number 1 • Winter 2023

Symposia

Trade Sanctions and International Relations
T. Clifton Morgan, Constantinos Syropoulos, and Yoto V. Yotov, "Economic Sanctions: Evolution, Consequences, and Challenges" 3
Marco Cipriani, Linda S. Goldberg, and Gabriele La Spada, "Financial Sanctions, SWIFT, and the Architecture of the International Payment System" 31

Monetary Policy
Anil K Kashyap and Jeremy C. Stein, "Monetary Policy When the Central Bank Shapes Financial-Market Sentiment" 53
Michael D. Bauer, Ben S. Bernanke, and Eric Milstein, "Risk Appetite and the Risk-Taking Channel of Monetary Policy" 77
Alan S. Blinder, "Landings, Soft and Hard: The Federal Reserve, 1965–2022" 101
Alisdair McKay and Christian K. Wolf, "Monetary Policy and Inequality" 121

Hispanic Americans
José Fernandez, Mónica García-Pérez, and Sandra Orozco-Aleman, "Unraveling the Hispanic Health Paradox" 145
Francisca M. Antman, Brian Duncan, and Stephen J. Trejo, "Hispanic Americans in the Labor Market: Patterns over Time and across Generations" 169
Gordon Hanson, Pia Orrenius, and Madeline Zavodny, "US Immigration from Latin America in Historical Perspective" 199

Article
Andrew Atkeson and Gita Gopinath, "Oleg Itskhoki: 2022 John Bates Clark Medalist" 223

Feature
Timothy Taylor, "Recommendations for Further Reading" 241

Statement of Purpose

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

Policy on Data Availability

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

Policy on Disclosure

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

Journal of Economic Perspectives—Volume 37, Number 1—Winter 2023—Pages 3–30

Economic Sanctions: Evolution, Consequences, and Challenges

T. Clifton Morgan, Constantinos Syropoulos, and Yoto V. Yotov

Many countries responded to Russia's invasion of Ukraine in February 2022 by imposing a wide range of economic sanctions on Russia and by escalating their use over time. These sanctions, which are among the most extensive ever imposed on a significant economic power, have stirred a great deal of interest in sanctions as an instrument of foreign policy and in what we know about when and how they work. We will define sanctions as "restrictive policy measures that one or more countries take to limit their relations with a target country in order to persuade that country to change its policies or to address potential violations of international norms and conventions" (Morgan, Bapat, and Krustev 2009; Syropoulos et al. 2022). Sanctions can serve many purposes and take many forms. Sanctions have been used to promote democracy, further human rights, combat terrorism and nuclear proliferation, destabilize regimes, and hasten the end of military conflicts. Sanctioning states commonly seek to curtail trade or foreign aid with the target state, but they can also restrict travel, freeze assets, and deny access to financial institutions by specified individuals or groups.

■ *T. Clifton Morgan is Albert Thomas Professor of Political Science, Rice University, Houston, Texas. Constantinos Syropoulos is Trustee Professor of International Economics, School of Economics, LeBow School of Business, Drexel University, Philadelphia, Pennsylvania. Yoto V. Yotov is Professor of Economics, School of Economics, LeBow College of Business, Drexel University, Philadelphia, Pennsylvania, and Research Professor, ifo Institute for Economic Research, Munich, Germany. Their email addresses are morgan@rice.edu, c.syropoulos@drexel.edu, and yotov@drexel.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.1.3.

This essay will focus on patterns of and lessons about economic sanctions since 1950. To aid our analysis, we will rely on data from the third release of the Global Sanctions Data Base (Syropoulos et al. 2022), covering 1,325 sanction cases during the period 1950–2022. A sanction case is defined for each year in which there is an active sanction imposed by a sanctioning state ("sender") on a sanctioned state ("target"), with senders and targets usually being individual countries or, especially in the case of senders, groups of countries, including international organizations (like the European Union or the United Nations).

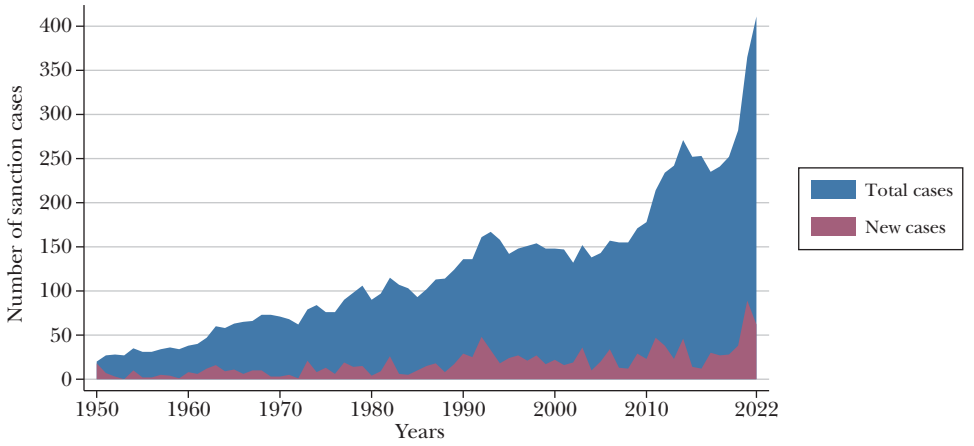
Consistent with the salient features and evolution of sanctions, the Global Sanctions Data Base identifies sanction cases on the basis of three distinct characteristics/dimensions related to type, sender, and objective. For example, cases are classified according to six sanction types: trade sanctions, financial sanctions, travel restrictions, arms sanctions, military assistance sanctions, and "other" sanctions, which do not fall into any of the main categories. There are nine sanction objectives, including changing policy, destabilizing regimes, resolving territorial conflicts, fighting terrorism, preventing war, ending war, restoring and promoting human rights, restoring and promoting democracy, and other objectives. Finally, for all sanctions that have been terminated/lifted, the Global Sanctions Data Base assigns a success score that corresponds to each sanction objective and varies from "total success" to "partial success", "settlement," and "failure."[1]

While, as we discuss below, there have been changes in the patterns of sanctions use over time, one trend stands out: there has been a phenomenal increase in the use of sanctions as an instrument of foreign policy since the end of World War II. Figure 1 highlights the steady increase in sanctions use over the past 70 years—there are roughly ten times as many active sanctions per year now as in 1950. Moreover, while a relatively small number of countries initiate the imposition of sanctions (the "primary senders"), often these sanctions trigger the imposition of countersanctions, or some other form of retaliatory response(s), with repercussions on nearly every state in the international system. In the next section, we trace some of the historical patterns in the use of sanctions. We will suggest that many of these patterns can be linked to evolving features of the international political and economic system.

Along with the rapid growth in the use of sanctions, research into sanctions processes has expanded in recent decades. One of our objectives in this essay is to identify some of what we know about the use of sanctions and about their effectiveness as instruments of foreign policy. An interesting aspect of this research is that

[1]The Global Sanctions Data Base is constructed from publicly available sources and cross-checked against other databases. Two prominent examples are the Hufbauer-Schott-Elliott sanctions database, (Hufbauer, Schott, and Elliott 1990; Hufbauer et al. 2007) and the Threat and Imposition of Economic Sanctions (TIES) database (Morgan, Bapat, and Krustev 2009; Morgan, Bapat, and Kobayashi 2021) which also includes sanction threats. For definitions, examples, and additional details on the Global Sanctions Data Base, see Felbermayr et al. (2020), Kirilakha et al. (2021), and Syropoulos et al. (2022). The Global Sanctions Data Base is freely available to researchers, although access must be requested by email at GSDB@drexel.edu. Details are available at http://www.globalsanctionsdatabase.com/.

Figure 1
Evolution of Sanction Cases, 1950–2022



Source: The figure is produced by the authors with the full sample from the third release of the Global Sanctions Data Base.
Note: This figure displays the evolution of existing and new sanction cases (of any type) over the period 1950–2022. The coverage stops before the middle of 2022.

economists often evaluate sanctions based on their economic effects, while political scientists often evaluate sanctions based on whether the ultimate goals of the sanctions are achieved. We suggest a need to bring these perspectives together if we are to continue improving our understanding.

One pattern that emerges from our investigation into the changing patterns in the use of sanctions is that contemporary sanctions policies are quite different from those adopted in earlier decades. Recent uses emphasize, to a much greater extent, the necessity of designing sanctions to target key individuals, companies, or sectors (for example, "smart" or "targeted" sanctions including financial and travel sanctions) rather than using sanctions as a blunt instrument designed to harm the entire target nation (for example, trade sanctions). Our theoretical and empirical understanding of sanctions has not kept up with these changes. Finally, an important contribution of this essay is that, in the conclusion, we lay out some of the new questions being raised by contemporary sanctions policies that need to be addressed in future research.

Evolution of Economic Sanctions, 1950–2019

Although our focus in this section is on the use of economic sanctions since 1950 and on the possible lessons we can draw from studying their patterns over time, we would be remiss if we did not note that debates over the merits of economic sanctions as an instrument of foreign policy have existed for at least 2,500 years.

One of the first (if not the first) recorded instance of economic sanctions can be found in the Megarian Decree (circa BCE 432), by which Athens barred trade with Megara and denied the Megarians access to Athenian ports. Ostensibly, this action was taken as a diplomatic measure against the Megarians for having cultivated Athenian land and killing an Athenian herald. The Megarian Decree is viewed by many as a trigger, if not a major cause, of the Peloponnesian War that followed. One view is that these sanctions were a way of imposing costs on the Megarians without having to resort to war, in which case the sanctions failed to achieve their goal. On the other side, many believe that Pericles persuaded the Athenians to adopt the Megarian Decree precisely because he intended to foment war (Kagan 1969; MacDonald 1983), in which case they would count as a success in achieving a dubious goal.

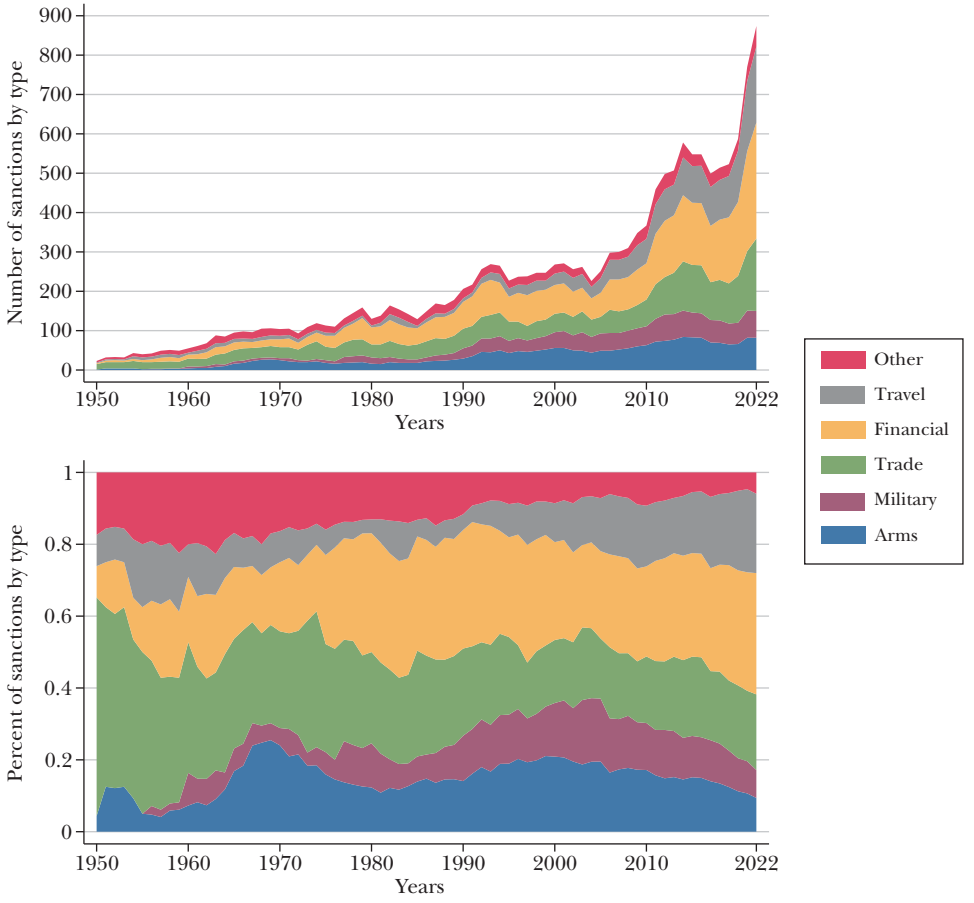
In the aftermath of World War I, there was broad interest in the use of economic sanctions as an alternative to war. However, the two best-known examples of sanctions from this time had questionable results as well. In one example, the League of Nations imposed sanctions on Italy in 1935 in response to its invasion of the Abyssinia region of Ethiopia. At that time, however, Italy was viewed as a counterweight for Nazi expansionism in Germany, which made countries like Great Britain and France unwilling to administer such sanctions. Ultimately, these sanctions are deemed to have been a colossal failure that undermined the standing of the League. In the second example, the United States imposed severe trade restrictions on Japan to discourage Japanese military conquests in East Asia. Instead, these sanctions contributed to the Japanese decision to widen the war by attacking Pearl Harbor in 1941 (Boudreau 1997; Hosoya 1968; Russett 1967)—another case of sanctions preceding military conflict, rather than being substituted for it.

The dramatic increase in the use of sanctions in recent decades is the result of a steady acceleration that began after World War II. There have also been significant changes in the purposes to which sanctions have been applied, in the types of sanctions used, and in the international actors who have imposed sanctions. Here, we outline some of these changes and associate them with significant events in the international political and economic system. To organize the discussion, we identify four eras, but this taxonomy is largely for convenience. We do not intend to imply that there were precise dates and sharp changes in sanctions usage.

Early Cold War: 1950–1975

During this period, close to one-third of the implemented sanctions were imposed by the United States acting unilaterally. Because the United States also played a key role in many of the sanctions imposed by the United Nations and by ad hoc multilateral coalitions, it was, by far, the most prolific sanctioner and thus largely responsible for the expansion in the use of sanctions. Figure 2, which shows the type of sanctions imposed, and Figure 3, which displays the purpose of sanctions, guide our discussion of the evolution of sanctions. About 60 percent of the sanctions in the early 1950s were trade and arms embargoes. These sanctions were most commonly applied to destabilize political regimes or to influence the

Figure 2
Evolution of Sanctions by Type, 1950–2022

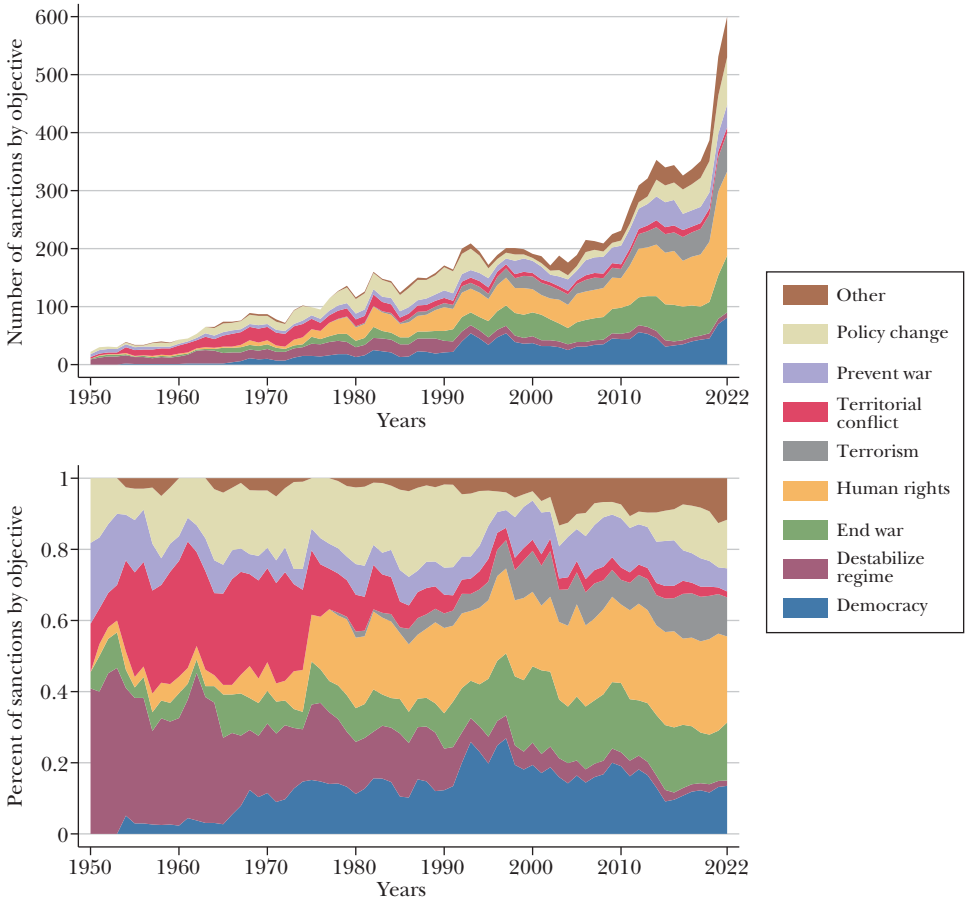


Source: The figure is produced by the authors with the full sample from the third release of the Global Sanctions Data Base.
Note: This figure displays the evolution of sanctions depending on their type over the period 1950–2022. The coverage stops before the middle of 2022. The top panel depicts the evolution of the number of sanctions in levels, while the bottom panel presents the same relationship as percentage shares. The range on the y-axis of this figure is longer as compared to the range in Figure 1 because some sanction cases include more than one type of sanction. We refer the reader to Felbermayr et al. (2020) and Syropoulos et al. (2022) for definitions and examples for the alternative types of sanctions.

direction of military conflict (either by pressuring combatants to end fighting or in support of senders' territorial claims).

During this period, the recent experience of World War II reinforced the desire for an alternative to military force. Hirschman's seminal *National Power and the Structure of International Trade* (1945) provided the theoretical basis for believing sanctions could provide that alternative. Hirschman argued that trade significantly

Figure 3
Evolution of Sanctions by Objective, 1950–2022



Source: The figure is produced by the authors with the full sample from the third release of the Global Sanctions Data Base.
Note: This figure helps visualize the evolution of sanctions depending on their objective over the period 1950–2022. The coverage stops before the middle of 2022. The top panel depicts the evolution of the number of sanctions in levels, while the bottom panel presents the same relationship as percentage shares. The range on the y-axis of this figure is longer as compared to the range in Figure 1 because some sanction cases include more than one objective. We refer the reader to Felbermayr et al. (2020) and Syropoulos et al. (2022) for definitions and examples for the different sanction objectives.

improves economic well-being for all, but that it also creates asymmetrical power relationships; that is, when states are interdependent through commerce, the less dependent nation may use restrictive trade policies to enhance its power by gaining leverage in disputes over other issues. Many high-income nations had a substantial share of their productive capacity destroyed by World War II, with the US economy

being a notable exception. In addition, soon after the end of World War II, the Cold War between the Soviet Union and the United States became the defining feature of international politics, and the United States frequently used sanctions in support of its Cold War policies (Barber 1979).

Considering these factors, it is not surprising that sanctions in this period were largely a tool used by the United States, that trade sanctions and arms sanctions remained the "weapons" of choice, and that sanctions were often used in efforts to bring about regime change. The US sanctions on Cuba, in effect to this day, are a prominent example. The United States also frequently imposed sanctions on the Soviet bloc countries of Eastern Europe; it resisted allowing China to integrate into the global economy; and it often sanctioned countries that it perceived to be "going communist."

Late Cold War: 1975–1990

By the early 1970s, Western Europe and Japan had largely recovered from World War II and began to challenge the economic hegemony of the United States (Blum 2003; Mastanduno 2019). What was then called the European Economic Community (EEC) was beginning to realize its promise of allowing its members to act with a single voice. Although the United States remained the single largest economy in the global system, a number of economic "shocks" (such as the unilateral decision by President Richard Nixon to allow the exchange rate of the US dollar to float and the OPEC oil embargo of 1974–1975, which led to higher oil prices that buttressed Soviet power and likely to increased aggressiveness elsewhere in the world) highlighted the United States's growing vulnerabilities. Politically, the United States was weakened by its long and unsuccessful war in Vietnam. The US public was less willing to use military force and the United States had lost some of its standing as a proponent of democratic values and human rights (Eichenberg 2005; Jentleson and Britton 1998). In much of the world, coercive dictatorships and military juntas had replaced earlier colonial governments and fledgling democracies. During this time, the Cold War was in a period of détente between the United States and the Soviet Union, the most notable exception being in the aftermath of the Soviet invasion of Afghanistan—which was met with significant economic sanctions. Furthermore, a number of guerrilla organizations (such as the Baader-Meinhof gang in West Germany, the Red Brigades in Italy, and the Symbionese Liberation Army in the United States) began using terrorist tactics.

The effect of these factors can be seen in the changing patterns of sanctions use in Figures 2 and 3. The acceleration in the use of sanctions continued, and the United States remained the single most frequent sender. Notably, the International Emergency Economic Powers Act of 1977 provided the US president with broad authority to regulate a variety of economic transactions following a declaration of national emergency. Indeed, this law became the "all-purpose" statute for US sanctions, and the frequent use of economic sanctions by the United States was at odds with customary international law at the time (Hufbauer 1998).

However, the Europeans also began to emerge during this period as coordinated sanctioners.

Throughout the period, trade sanctions continued to be used consistently, but this became a smaller proportion of the overall total—a steady increase in the use of financial and military types of sanctions had already begun. While the use of sanctions for regime destabilization continued at a fairly constant rate, there were the beginnings of a dramatic increase in the use of sanctions for the purpose of protecting human rights. By the end of the period, sanctions were being used extensively to combat international terrorism (see also Elliott and Hufbauer 1999; Hufbauer and Moll 2007; Choi and Luo 2013; Zanchetta 2016).

Post-Cold War: 1990–2000

In the 1990s, a great hope emerged that the global polity was on the brink of a "New World Order" in which democracy and liberal economic relationships would spread, interstate war would be (largely) a thing of the past, and international organizations would manage conflict and structure cooperation (Barnett 1997). In the early 1990s, the Cold War came to an abrupt end with the collapse of the Soviet Union, and a wave of democratization swept much of the globe. Iraq's invasion of Kuwait in August 1990 was met by an unprecedented level of international cooperation, which increased the esteem and strengthened the position of multilateralism, especially through international organizations such as the United Nations. The United States adopted the policy of fully incorporating China in the international economic system (Jacobson and Oksenberg 1990). On the economic side, the 1992 Treaty of Maastricht, which created the European Union, set Europe on the path of even greater economic coordination. In 1995, the World Trade Organization superseded the 1947 General Agreement on Tariffs and Trade (GATT), regularizing trade relationships even further, but also making it more difficult to apply trade sanctions (Charnovitz 2001; Mitchell 2017).

Of course, narratives are never this simple. In retrospect, seeds of opposition were also planted. International terrorist organizations, notably Al Qaeda, began to gather strength. At the same time, there was backlash and protests against globalization. These forces would become especially prominent after 2000.

During the 1990s, the frequency with which sanctions were used remained fairly constant at historically high levels and the 1990s became known as the "sanctions decade" (Cortright and Lopez 2000), although they were but a hint of what was to come. Some changes can be noted, however. Most obviously, there was no expansion (and perhaps a contraction) in the overall use of trade sanctions. However, there was an increased use of financial sanctions and, especially, sanctions involving arms transfers. In terms of the issues over which sanctions were imposed, this period saw a significant reduction in the use of sanctions to bring about regime change and significant increases in the use of sanctions to promote democracy and human rights. Most notably, while many sanctions were still imposed unilaterally, the proportion of sanctions imposed multilaterally increased substantially, with the United Nations and the European Union greatly expanding their use.

Post-9/11: 2001–Present

The terrorist attacks of September 11, 2001, in the United States unleashed two decades of war. This period also witnessed a turn away from democratization and a turn toward nationalism. More recently, the world has also experienced major economic upheaval following the financial crisis of 2008 and the global pandemic. Once again, we are experiencing an interstate war in Europe, which has triggered the most substantial sanctions ever imposed on a relatively strong economic power.

As shown in Figure 1, the increase in the use of sanctions for a decade or so after 2001 was unprecedented.[2] This increase has been largely in the form of financial, travel, and other sanctions targeting specific individuals and enterprises. While the use of sanctions to promote human rights has continued to rise, their use to promote democracy has leveled off. Sanctions have rarely been used to bring about regime change in this period, but their use to combat international terrorism has increased substantially.

Several factors have been instrumental in fostering these changes. First, the increased reliance on targeted sanctions followed on the heels of theoretical advances suggesting they should be more effective (Cortright and Lopez 2002; McGillivray and Stam 2004; Bapat et al. 2013; Peksen 2019). Second, the United States enacted a number of changes to its laws that make it much easier to impose and enforce financial sanctions (Hufbauer and Moll 2007). In its effort to undermine the activities of terrorist organizations, the United States now requires financial institutions to track and report financial transactions. This has induced the United States to target sanctions at specific firms and individuals and to push much of the enforcement of these sanctions onto financial institutions. The United States has also asserted extraterritorial jurisdiction over foreign entities that conduct business in US dollars and/or route payments through US financial institutions (Hufbauer and Jung 2020). Additionally, the United States has imposed secondary extraterritorial sanctions; that is, sanctions on enterprises, including foreign companies, for doing business with entities on its sanctions list. Third, the extraordinary advances in information technology have made it possible to process information on the vast number of financial transactions that occur on a daily basis. While we have theoretical reasons to believe these changes in sanctions design should make sanctions more effective and reduce the harm on innocent civilians, as we discuss in the next section, we do not yet have the ability to assess systematically whether this promise is being realized.

[2]We note that the data suggest there has been a sharp increase in sanctions during the past three years. It is too soon for us to draw conclusions about, or from, this. In addition to the many sanctions normally imposed by the European Union and the United States, a large number of the recent cases involve Great Britain (acting as an individual sender due to its separation from the European Union), and Russia and Belarus (as targets due to the use of military force in Ukraine).

The Economic Impact and Political Success of Sanctions

Almost all of the research on sanctions has focused on questions regarding their effectiveness as an instrument of foreign policy.[3] However, economists have tended to interpret "effectiveness" in terms of the economic damage that sanctions cause, while political scientists have considered sanctions "effective" only if they achieve their declared political objectives. The dividing line between sanction "effectiveness," "success," and "impact" is further blurred by the possibility that some issues are "fake"—in the sense that the overt political demands are intended only to provide cover for another agenda (Hufbauer and Jung 2020). In addition, it is difficult to determine "success" when political objectives shift over time. In a recent example, sanctions were first imposed on Russia to dissuade it from invading Ukraine in 2022. But when sanctions failed to achieve that purpose, new purposes emerged: to punish Russia for its invasion; to provide indirect support to Ukraine to fight back against the invasion and induce Russia to end the war; and, in the words of US Secretary of Defense Lloyd Austin, to weaken Russia "to the degree that it can't do the kinds of things that it has done in invading Ukraine" (as reported by Ryan and Timsit 2022).

To minimize ambiguity, we maintain the following distinction in terminology: (1) "economic impact" (or just "impact") will refer to the economic damage and costs of sanctions, and (2) "political success" (or just "success") will refer to sanctions' capability to achieve their announced political objectives. At the most basic level, our understanding of sanctions processes suggests that these notions of effectiveness are connected and that both are important: sanctions are intended to impose economic costs on the target who is supposed to be persuaded to alter its behavior in order to avoid paying those costs. However, it seems unwise to presume either that significant impact leads to success or that lack of success implies little previous impact. Furthermore, advances in our understanding of sanctions processes require that we account for the interconnections between these disciplinary differences.

Economic Impact of Sanctions

Broadly speaking, economists have concluded that sanctions produce significant economic impacts. These sanction effects can arise directly through their impact on the target country. They can also arise indirectly through reciprocal sanctions aimed back at the primary sanctioner, as well as through third party nations. We now consider each of these three main actors in turn.

The primary focus of most of the empirical literature on sanctions has been on how sanctions affect target states. To quantify these effects, scholars have used different datasets and a variety of econometric methods. Four general conclusions

[3] Interestingly, almost none of the research on the use of military force has addressed the question of whether states that use military force actually achieve their objectives. Instead, researchers in this area have been more interested in determining the conditions that lead to the deployment of force.

have emerged in the related literature.[4] First, the impact of sanctions on various economic agents (firms and individuals), sectors, and specific activities in target states has been negative and significant. Second, economic sanctions have had strong negative effects on the overall performance of the sanctioned states—including trade, foreign direct investment, growth, poverty, and political stability. Third, the effects of sanctions on economic development, trade flows, foreign direct investment, and growth are long-lasting and often persist even after sanctions are lifted. Fourth, the effects of sanctions can be very heterogeneous depending on their type (for example, trade versus financial sanctions or complete versus partial sanctions), on whether they are imposed unilaterally or multilaterally (for example, UN versus US sanctions), and on the specifics of individual cases. Thus, models that impose common sanction effects across the various sanction dimensions noted earlier often mask the presence of very significant heterogeneity.

Often, to mitigate the adverse effects of sanctions, sanctioned states take actions to redirect their international trade and investment flows toward third countries, "shield" certain economic agents (often large firms perceived as "vital" to national interests), form alliances with "friendly" third countries and, as is the case with Russia in 2022, retaliate against their sanctioners. It is difficult to identify, based on official documents, whether certain sanctions are retaliatory. However, an inspection of the Global Sanctions Data Base reveals that the number of retaliatory sanctions is very small: often, they are best regarded as symbolic statements.

From a methodological perspective, many studies aiming to estimate the impact of sanctions on target states face problems of endogeneity in the following sense: events that instigate the sanctioning of target countries—for example, civil or interstate conflicts or violations of human rights—may also shape the economic effects we observe. Surprisingly, much of the extant literature has bypassed this issue. Gutmann, Neuenkirch, and Neumeier (2020) and Kwon, Syropoulos, and Yotov (2022a) are recent exceptions. Capitalizing on certain dimensions of sanctions, these studies have addressed the issue of endogeneity; for example, by considering flexible instruments related to laws and regulations in sanctioning states that are independent of events in sanctioned states. It is important for future studies of the effects of sanctions to recognize the endogeneity problem and tackle it directly, either with existing methods or with new strategies.

So far, interest among academics in the impact of economic sanctions on *senders* has been limited, perhaps because this impact tends to be relatively small. A possible explanation for relative disinterest may rest in the fact that the economies of most sanctioning states are considerably larger than the economies of the targeted states, which tends to weaken *bilateral* economic dependence. Moreover, in most cases, senders that are threatened with reciprocal countersanctions can divert economic activity toward third, nonsanctioned states. In addition, senders may design and/or

[4]We refer the reader to Felbermayr et al. (2021) and van Bergeijk (2021) for a review and a compilation of recent applied work on the economic impact of sanctions, respectively. For a more extensive set of references, see Morgan, Syropoulos, and Yotov (2022).

implement sanctions with a view toward minimizing, or at least mitigating, the possibly adverse impact of sanctions on their constituencies. The current sanctions on Russia are a prominent example: some countries decided not to impose sanctions on it, and others failed to enforce their declared sanctions fully. Naturally, sender efforts to minimize their own costs raise deep questions about the effectiveness, enforcement, and credibility of their sanctions policies (Lektzian and Sprecher 2007).

Despite the eagerness and ability of senders to minimize the negative impact of sanctions on their own economies, recent quantitative analyses provide some evidence for the presence of such negative effects due to significantly decreased economic activities between senders and targets (Besedeš, Goldbach, and Nitsch 2021). Nonetheless, these effects do not translate into a significant impact on senders due to several factors, including the intensified economic relationships among sanctioning and nonsanctioned countries (often called "trade diversion"), the disproportionate size between the primary sanctioners and targeted states, and the possible backsliding by some sender countries in a sanctioning coalition. Thus, consistent with the earlier literature (for example, Farmer 2000), recent evidence suggests that the impact of sanctions on sender states tends to be small and short-lived. However, this does not necessarily imply that the costs of sanctions to senders would be small if the targets are economically large and powerful.

Notwithstanding the limited impact of sanctions on senders, we see several promising directions for future work in this area. First, from a methodological perspective, the ability of senders to select an optimal mix of sanction tools and to design sanctions in a way that maximizes the economic damage on targets while minimizing the cost on the senders should be a key feature of theoretical models on sanctions. A second direction that may be interesting to explore, both theoretically and empirically, is related to the possibility that senders may issue "fake" sanctions based on political pronouncements aiming to camouflage their economic motives. Thus, the imposition of sanctions may be intended to provide gains for the sender rather than to fulfill the declared political objectives of sanctioning. This story is also consistent with the notion that sanctions may be issued to serve the interests of specific interest groups (Kaempfer and Lowenberg 2007). Third, the cost of sanctions among members in coalitions of senders may be shared disproportionately. This suggests a role for the adoption of reliable redistributional mechanisms within sender coalitions aimed at sharing the burden of sanctions, with implications for improvements in the design, implementation, and effectiveness of multilateral sanctions.

In addition to affecting senders and targets, sanctions may also affect third countries. Although these effects have been examined by policymakers and covered extensively in the media, they have attracted relatively little attention in the academic literature. One can identify two distinct and opposing channels through which sanctions may affect third countries: (1) the "extraterritorial" channel, which is a direct channel that normally transmits an adverse effect on third countries; and (2) the "general equilibrium" channel, which is an indirect channel through which sanctions generate (usually) positive effects on third countries.

The intuition behind the general equilibrium effects of sanctions and their impact on third countries is familiar and easy to understand. Economic activities that are disrupted by sanctions can intensify commercial, financial, and other relationships with third countries that serve as a substitute for forgone business opportunities. For example, after the imposition of sanctions on Russia in 2022, imports of Russian oil by India and China soared. These effects can be quantified, because the economics literature has developed tools to capture such general equilibrium effects (Haidar 2017; Besedeš, Goldbach, and Nitsch 2021).

Usually, the general equilibrium effects of sanctions on individual third countries are small because the diverted activity is distributed among numerous nonsanctioned states. Some suspect, however, that the general equilibrium impact on a target could be large if it accumulates across all nonsanctioned states, thereby mitigating the intended sanction costs from the primary sanction effects. This is one of the main motives for "extraterritorial" sanctions, to which we now turn.

We define "extraterritorial" (or "secondary") sanctions broadly as penalties on individuals, companies, organizations, and other entities from nonsanctioned countries due to their engagement in activities like trade, investment, or other business activities with a sanctioned state.[5] Prominent recent examples of such sanctions are the US threats and actions toward the German companies that were involved in the construction of the Nord Stream 2 natural gas pipeline running from Russia to Germany. Extraterritorial sanctions and their effects have been addressed in the popular press and in policy reports. Ample anecdotal evidence suggests that these effects exert a direct negative impact on nonsanctioned countries, which often view these actions as "forced" and/or "illegal" (Meyer 2009; European Commission 2021).

Despite the attention paid to the extraterritorial effects of sanctions by the media and policy analysts, the related academic literature is relatively scarce. Moreover, most of the existing studies are descriptive and offer limited qualitative evidence for the extraterritorial sanction effects (Gordon 2016). Several recent theoretical and quantitative studies offer evidence that extraterritorial sanctions cause significant additional economic damage to targeted countries, thereby contributing to the political success of sanctions (Han 2021; Kwon, Syropoulos, and Yotov 2022b). Nevertheless, in light of the importance of extraterritorial sanctions to sound policymaking and their role as a salient determinant of sanctions success (Early 2021), more work is needed to quantify their effects.

Overall, economists have concluded that economic sanctions can have significant, predictable, and often long-lasting effects on targets. Given our general

[5] In the popular media, "secondary sanctions" is often used as a synonym for "extraterritorial sanctions." However, there are important differences between these terms: "secondary" sanctions are aimed at entities that are not directly related to the sanctioning states, while "extraterritorial" sanctions are aimed at entities that are affiliated with the sanctioners but operate in nonsanctioned countries. In practice, the line between these terms is often blurred. To simplify exposition, we use the term "extraterritorial" sanctions in a broad sense to include sanctions on sender entities that operate abroad as well as sanctions on nonsender entities.

intuition regarding sanctions processes, this might lead us to expect sanctions to be an effective instrument of policy. Yet that is not what political scientists have generally concluded.

Political Success of Sanctions

Political scientists have long debated on whether sanctions "work" in the sense of achieving their stated goals. Early research was largely focused on prominent cases, such as the US sanctions on Cuba or the League of Nations sanctions on Italy mentioned earlier, and generally came to the conclusion that sanctions do not bring about significant changes in target state policies (Galtung 1967; Hoffmann 1967; Doxey 1972). However, it was quickly recognized that this work suffered from a severe selection bias—the reason that the cases under study were "prominent" was precisely because they failed. Early statistical analyses based on the well-known Hufbauer, Schott, and Elliott (1990) dataset indicated that sanctions achieve their political objectives in about one-fourth to one-third of the cases.

Much of the research into sanctions conducted by political scientists has focused on a puzzle: if sanctions seldom "work," then why do they continue to be applied, and at an increasing rate? Several broad approaches have been taken to address this puzzle. First, some argued that although sanctions seem ineffective at achieving their stated objectives, they may be relatively effective in achieving their "true" objectives. For example, some sanctions may aim to support domestic interests (Kaempfer and Lowenberg 2007), while others may aim to serve symbolic (Lindsay 1986) or signaling (Schwebach 2000) purposes.

Second, several theoretical arguments suggest that sanctions should not be expected to achieve their objectives except under very specific conditions (Morgan and Schwebach 1997). For example, Wagner (1988) posited that if we applied bargaining theory to the agreements that produced the economic exchanges that sanctions disrupt, we would conclude that sanctions should, in most cases, harm the sender as much as they harm the target. In other words, the leverage provided by sanctions cuts both ways. Morgan, Bapat, and Kobayashi (2021) offer evidence suggesting that sanctions often "work" at the threat stage; consequently, successful sanctions might not actually be imposed.[6]

Third, even in their worst light, sanctions have been shown to be effective in a modest fraction of cases. Even a 25 percent success rate for sanctions may be considerably higher than doing nothing, and the costs may be substantially lower than other alternatives, like overt military interventions. Perhaps the "sanctions glass" should be viewed as one-quarter full, not three-quarters empty.

Finally, it may be possible to identify specific factors that lead to increases in the costs that sanctions impose on targets and thus to determine when sanctions have been ineffective and how to make them more likely to be effective. For example,

[6]This argument led to the development of the Threat and Imposition of Economic Sanctions (TIES) data set. For data and details on the imposition of sanction threats, we refer the reader to Morgan, Bapat, and Krustev (2009) and Morgan, Bapat, and Kobayashi (2014; 2021).

Attia, Grauvogel, and von Soest (2020) suggest that poor economic health and high political volatility in targets are important determinants of sanctions success. Others have found that the extent of the interrupted economic relationship is a significant factor in sanctions success (Bapat et al. 2013). Moreover, the availability of sanctions busters, or "Black Knights," can enable targets to avoid significant costs (Early 2011). Relatedly, multilateral sanctions, especially when imposed under the auspices of an international organization, increase target costs relatively more than unilateral sanctions (Martin 1992; Bapat and Morgan 2009; Early 2021). Finally, sanctions are more likely to be effective when imposed on democracies than when imposed on autocracies, because democratic governments are more susceptible to costs felt by their populaces (Allen 2008; Lektzian and Souva 2007).

However, all of the above findings appear quite sensitive to model specification (Bapat et al. 2013). Indeed, given alternative models and specifications, the weight of the evidence might even turn against these findings (Demena et al. 2021).

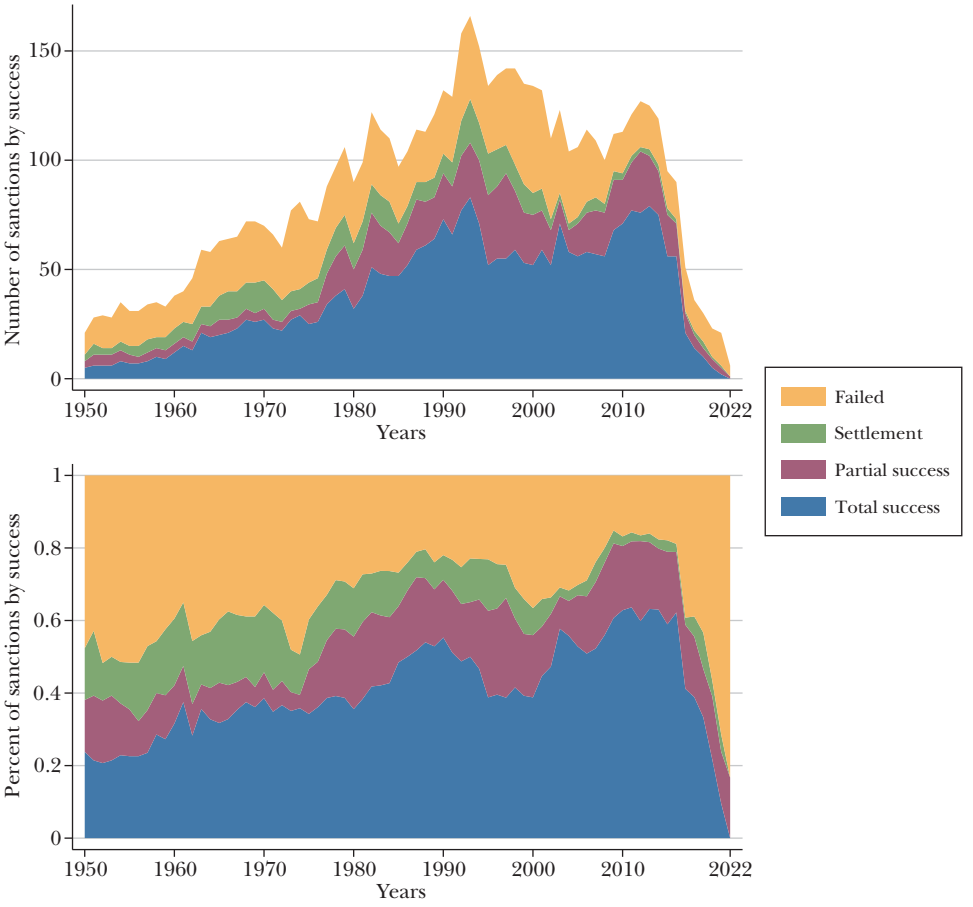
One important issue is that studies of the success of sanctions face endogeneity problems, just like studies of their economic impact. For example, senders control aspects of the design of sanctions including what sanctions are imposed (say, trade bans, asset freezes, and retraction of foreign aid) and the issues or demands to be met for lifting the sanctions (for example, ending war or improving human rights). Given that these choices are intertwined, it might be difficult to separate the effect of sanctions from the intractability of the underlying issues. Some studies suggest that endogeneity effects are serious in this body of work. For example, Morgan (1995) showed with multinational data over several decades that a threat of minor sanctions could be more effective than a threat of major sanctions if the accompanying demands were properly scaled. Similarly, Biersteker and van Bergeijk (2015) confirm that the success rate is higher for sanctions with narrowly defined goals and for sanctions that are accompanied by additional policy instruments.

Simple analysis based on the Global Sanctions Data Base provides additional insight. This data recognizes that sanctioners often have multiple goals and assigns a separate success score (varying from "total success" to "partial success," "settlement," and "failure") to each sanction objective.[7] This classification is based on information from official government statements or indirect confirmations in international press announcements, which indicate whether sanction objectives have been achieved.[8] The top panel of Figure 4 traces the evolution over time of the number of sanctions by success, while the bottom panel transforms these same numbers to percentage terms. Since some sanctions include more than one

[7]The Global Sanctions Data Base classifies an outcome as "settlement" when the sanctioning and sanctioned parties agree to settle a conflict with negotiations. The final success of the initial policy objective remained unclear, however, after sanctions were lifted. We refer the reader to Felbermayr et al. (2020) for detailed definitions and examples of the different success categories.

[8]Admittedly, there is still an element of subjectivity, especially for the two middle categories. Hufbauer and Schott (1985), Hufbauer, Schott, and Elliott (1990), and Hufbauer et al. (2007) offer more detailed classifications of sanction success for a small fraction of the cases included in the Global Sanctions Data Base.

Figure 4
Evolution of Sanctions by Success, 1950–2022



Source: The figure is produced by the authors with data from the third release of the Global Sanctions Data Base.
Notes: This figure illustrates the evolution of sanctions depending on the success of reaching their individual political objectives, 1950–2022. The coverage stops before the middle of 2022. Since some sanctions include more than one objective, success is defined for each individual objective. Ongoing sanctions are not included in the data used to construct these figures. The top panel depicts the evolution of the number of sanctions in levels, while the bottom panel presents the same relationship as percentage shares.

objective, success is defined for each individual objective. Several patterns emerge from this figure.

First, in the top panel of Figure 4, we see that the number of sanctions whose objectives are defined as "successfully met" increased steadily until the early 1990s, when it reached a peak. It reached another peak around 2013 and has fallen since then. It seems that the drop in the number of successful sanctions is particularly

pronounced during harder economic times (like the recessions in the early and late 2000s or the recent COVID period). Second, consistent with the earlier literature that argued sanctions do not work (Galtung 1967; Pape 1997; Hufbauer et al. 2007), the bottom panel of the figure reveals that the share of "successful" sanctions is relatively small—about 42 percent on average. Third, despite some slowdowns (for example, during the late 1990s and early 2000s), the share of successful sanctions has increased continuously over time, until 2016, suggesting that sanctions have become more effective in recent years. One possible explanation is that the effectiveness of sanctions has improved due to significant learning effects in their application (Early 2021). Finally, the share of sanctions that successfully reached their objectives decreased dramatically in recent years, a period coinciding with the Trump presidency and COVID.

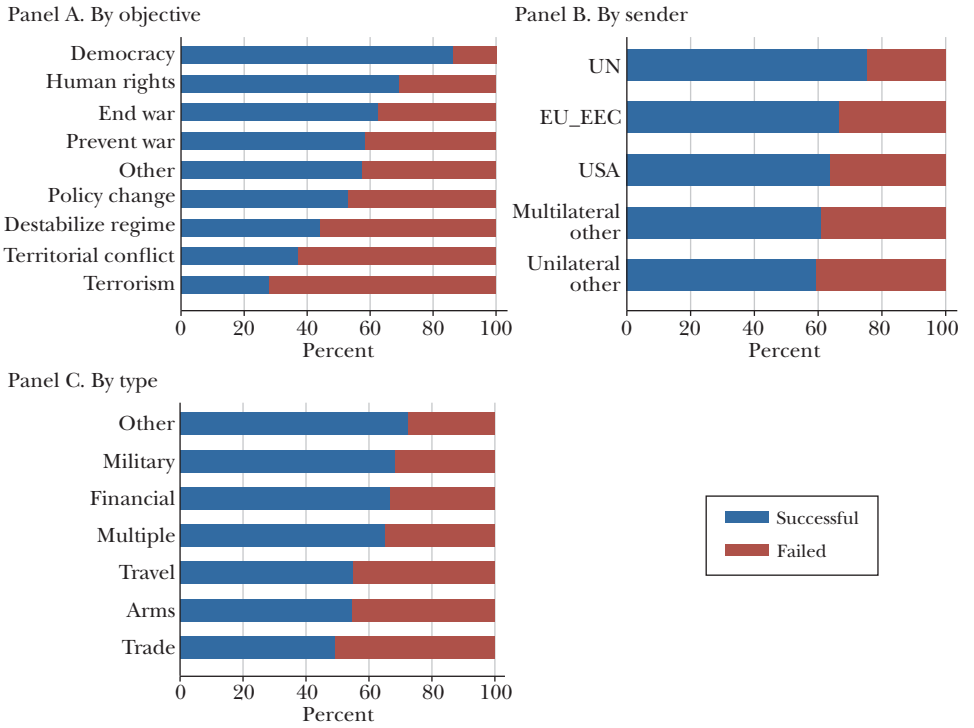
Figure 5 depicts the success of sanctions across three dimensions—objective, sender identity, and type—and suggests three lessons. To simplify the analysis of sanctions success by objective, we transform the four success categories into a bilateral success index, which is defined as "success" if the corresponding success category in the Global Sanctions Data Base is "total success" or "partial success," and as "failure" if the corresponding success category is "settlement" or "failure." The top panel of Figure 5 suggests that the most successful sanctions are those that aim to improve democracy and human rights, while the least successful are those aiming at regime change, territorial disputes, and terrorism. The category "Other," which includes objectives that are not prominent enough to form a separate group, is also among the relatively successful categories. The explanation for their seeming success is that sanctions in the "Other" category most often aim to achieve very specific and tangible objectives (like ending drug trafficking, releasing imprisoned citizens, and fighting corruption).

As noted earlier, the Global Sanctions Data Base assigns a unique success category to each sanction objective. However, because some sanctions include more than one objective, it is not possible to assign unique success categories by sanction case, sender, or type of sanction. Therefore, to analyze sanction success by sender and by type of sanction, we construct a simple composite success score index. Specifically, we first assign values ranging from one to four to the original categories of failure, settlement, partial success, and total success, respectively. Then, for each sender or sanction type, we take the average across the corresponding success scores.[9]

To obtain the results in panel B of Figure 5, we distinguish between multilateral and unilateral sanctions based on whether there are one or more senders involved in a particular case. Moreover, given their prominence as sanctioners, we isolate US sanctions from other unilateral sanctions, EU sanctions, UN sanctions, and other

[9] The vast majority of sanction cases in the Global Sanctions Data Base (95 percent) only have one or two objectives. Moreover, in cases with two objectives, only a small subset of them (less than 10 percent) has success categories running in opposite directions; for example, partial success for one, but failure for the other. The results we present remain qualitatively intact if we drop all cases with success categories in the opposite direction.

Figure 5
The Determinants of Sanction Success



Source: The figure is produced by the authors with data from the Global Sanctions Data Base.
Notes: This figure displays the relationship between sanction success and various sanction characteristics. Panel A depicts the success of sanctions depending on their objectives and relies on a binary success score measure as described in the main text. Panels B and C describe sanction success by sender and by type of sanction, respectively. See text for further details on the success score indexes used in panels B and C.

multilateral sanctions. Panel B unveils several intuitive findings. Sanctions imposed multilaterally are more likely to be successful than sanctions imposed unilaterally, especially when an international organization is involved.[10] Moreover, sanctions

[10]The importance of multilateral organizations in the success of sanctions has been recognized in the literature. Drezner (2000, p. 75) notes: "International organizations can turn fragile agreements to cooperate into a robust coalition by enforcing a previously agreed-on equilibrium . . . by acting as a coordinating mechanism for reassurance and information, enabling governments to resist domestic pressures, and providing side payments to increase the value of continued cooperation." Bapat and Morgan (2009, p. 1975) subject the most prominent arguments against the multilateral approach to empirical scrutiny and conclude that ". . . multilateral sanctions do appear to work more frequently than do unilateral sanctions [depending] . . . on whether an international institution is involved." The result that EU sanctions are less successful than UN sanctions is consistent with Besedeš, Goldbach, and Nitsch

imposed by the United States have been more successful than other unilateral sanctions.

Turning to the success of sanctions by sanction type, which are presented in panel C of Figure 5, we see that smart/targeted sanctions (like financial and travel sanctions) are more likely to succeed than trade sanctions.[11] Recall from our discussion above that the more successful types have become more frequently used over time, while the less successful types have become less so. This shift could account for much of the increased overall success rate of sanctions (and would be consistent with an explanation based on learning effects).

Overall, the lessons that emerge from the analysis in this section do not seem altogether consistent. Sanctions do seem to cause significant economic damage to the targets across various dimensions. However, although it seems intuitively clear that economic damage and costs to the target states should be key factors affecting the probability for sanctions success, there is no robust evidence for a clear causal link between economic costs and the political success of sanctions. Moreover, while recent trends suggest the presence of an improvement in sanction effectiveness, overall sanctions are still not perceived as particularly successful policy tools.

Of course, we can suggest some possible hypotheses for the apparent inconsistency in these conclusions. For example, perhaps sanctions do cause economic harm, but often this harm may not be sufficiently strong to lead to political success. In addition, sanctions enforcement is often lax, to put it mildly (Morgan and Bapat 2003; Bapat and Kwon 2015). Comprehensive or secondary sanctions can impose significant costs on countries that wish to avoid those costs; indeed, many individual states do not have the infrastructure or capacity to detect, investigate, or prosecute sophisticated economic activities aimed at circumventing sanctions. Moreover, efforts to act through the United Nations and improve enforcement can be blocked by adversely affected actors. These issues might be addressed by improved design of sanctions and by more widespread international cooperation on enforcement, but the political obstacles to these steps should not be underestimated.

It might also be the case that we should judge the effectiveness of a sanctions policy in its totality, rather than the effectiveness of specific impositions of sanctions one at a time. We know that some targets back down when threatened, but senders must be willing to implement sanctions to maintain the credibility of their threats (Morgan, Bapat, and Krustev 2009; Bapat et al. 2013). In this scenario, the imposition of unsuccessful sanctions in some cases might be the price to achieving success by threats of sanctions in other cases. Moreover, we cannot know exactly how often human rights have *not* been violated or how many nuclear weapons have *not* been

(2017), while the finding that EU sanctions are more successful than US sanctions is consistent with Weber and Schneider (2020).

[11] The category "Other," which appears on the top of panel C, includes a small number of sanctions (see also Figure 2). Usually, these are diplomatic sanctions. Often diplomacy is restored in a relatively short time and once the main objective is achieved, at least to some extent. For example, diplomacy related sanctions are commonly imposed on African countries by other African countries due to civil wars or military coups. Diplomacy stops during coups, but is restored soon after.

tested because of implicit sanction threats. We clearly need better tools to test for the presence of links between the economic impact and political success of sanctions as well as improved methods to tackle the complexity in the evaluation of the sanction effects.

Future Questions and Challenges

The sanctions imposed on Russia in response to its invasion of Ukraine in 2022 are unprecedented in their scope and impact. This is the first time a large number of members of the World Trade Organization have imposed expansive and substantial punitive economic measures against another member.[12] For example, Russia's "most favored nation" treatment was revoked on the basis of a rule within the treaty (General Aggreement on Tariffs and Trade Article XXI) that permits members of the World Trade Organization to use trade restrictions to protect their "essential security" interests. Additionally, the sanctions on Russia aim to undermine its access to credit and debt, to limit the mobility of influential Russians and their assets, and to curtail exports to Russia of luxury goods and dual-use technologies/products of potential use to the military. Given the possibilities of evasion by private interests, "backsliding" by coalition partners, and "backfilling" by potential "spoilers" (such as China and India, whose impact, given their sizes and prominence in global affairs, may be significant), the United States also took steps to discourage misbehavior within and outside the coalition—for example, with threats of secondary sanctions for violators and promises to help alleviate the economic stress on allies facing limited supplies and rising costs of energy.

Will these sanctions significantly harm the Russian economy? Will they induce Russia to end its military actions against Ukraine? Will they discourage Russia (and, more broadly, other countries) from using military means to pursue geo-economic objectives in the future? Experience suggests that the sanctions will likely produce high and long-lasting economic costs on Russia, the sanctioning states, and several third countries. Moreover, the magnitude of future costs due to these sanctions is highly uncertain and potentially very large. However, these costs are unlikely to induce Russia to end its invasion, at least directly—although, to the extent that they inhibit Russia from resupplying its military, it is conceivable that sanctions may contribute to a battlefield success on the part of Ukrainian military forces.

We wish to end by raising several broader questions about the modern application of sanctions that are illustrated by the Russia-Ukraine experience, but which also arise more generally from the continued evolution of sanctions. The single

[12] While there is a certain similarity between Iraq's invasion of Kuwait in 1990 and Russia's invasion of Ukraine, as well as in the Western economic response in both cases, there are important differences. First, the size of Iraq's territory, military, and geo-economic power pale in comparison to Russia's. Moreover, Iraq was not a WTO member while Russia is, which complicates the West's economic response in the current case. In the end, the extraordinarily comprehensive sanctions of the United Nations failed to curb Saddam Hussein's aggression and military intervention was deemed necessary.

biggest trend in sanctions in recent decades is the extent to which essentially all countries in the world have become involved in imposing sanctions—given that members of the United Nations can be viewed as participating in sanctions. As the use of sanctions has become more widespread and sanctions have been imposed by various coalitions (such as the United Nations, the European Union, the African Union, and so on), a group of states that are targeted extensively by nearly all other states has emerged. This group, which in recent years included North Korea, Iraq, Afghanistan, and Iran, must now be amended to include another, more powerful nation: Russia. This situation raises questions that are likely to significantly shape future research on sanctions.

As a starting point, is the expanding use of sanctions creating security/defense threats as significant as those generated by increases in the destructive power of military weaponry? US Secretary of the Treasury Janet Yellen (2022) recently warned: "Going forward, it will be increasingly difficult to separate economic issues from broader considerations of national interest, including national security." To improve our understanding of sanctions and their effects, it is imperative that we develop models that account for the interconnections between military and economic security concerns. Analysis based on such models could seek to capture Hirschmann's (1945) insight, for example, that trade can serve as an instrument of power and may provide a workable foundation for developing valuable insights on the relationship between sanctions, military strength, and geopolitical objectives.

A nascent research effort along these lines is currently underway. The existing literature suggests that contest-based models may be fruitfully employed for this problem. Important contributions to this literature include Powell (1993), who builds a "guns-versus-butter" model to study the emergence of peace as a Markov perfect equilibrium, and Skaperdas (1992), Hirshleifer (1995), and Grossman and Kim (1995), who highlight the importance of incomplete property rights and operationalize conflict in general-equilibrium settings. Emphasizing the links between trade and security, Garfinkel, Skaperdas, and Syropoulos (2015) explore the importance of interstate competition over insecure resources (like territory) and show how trade, through its impact on product and factor prices, may affect the intensity of conflict and welfare. In similar spirit, Garfinkel, Syropoulos, and Yotov (2020) study how international trade among "large" trading partners and their political affiliates ("enemies" or "friends") condition their incentives to arm. Even more recently, motivated by Hirschmann's (1945) contribution, Garfinkel, Syropoulos, and Zylkin (2022) construct a dynamic model that captures the link between the gains from trade—which can be directed into arming and saving—to show how insecurity, conflict expectations, and the distribution of resources affect power and the expected gains from trade. A noteworthy insight of this work is that larger economies tend to experience an erosion of their security because their gains from trade (and thus their relative incentives to arm) tend to be relatively small. In a similar vein, Camacho et al. (2022) construct a general equilibrium model that explains how national security considerations may undermine countries' willingness to share or adopt military, dual-use, and civilian-use technologies. By shedding light on the

links between economic and geopolitical interests, studies of this type may also help bridge the gap between the political science and economics scholarship on sanctions.

The rising importance of sanctions, with both economic and security consequences, also underscores the importance of understanding how targets respond to them. As a vivid example of the issues ultimately involved here, when the United States and Western Europe imposed sanctions on Russia in response to its annexation of Crimea in 2014, Russia responded with costly countersanctions which led many nations in the West to either remove their own sanctions or to undermine the coalition's sanctions through lax enforcement (Bapat and Kwon 2015). It is distinctly possible that Russian President Vladimir Putin expected a similar outcome prior to invading Ukraine in 2022. This scenario emphasizes a broader point: just as senders are experimenting with how to impose sanctions more effectively, targets are experimenting with how to respond.

Finally, as states seek to reduce their vulnerability to sanctions, what ramifications may arise for the international economic and political system? For example, when members of the World Trade Organization mix trade and security policies, as has been the case in the Russia-Ukraine crisis, the survival of WTO and the rules-based approach to policymaking may be at risk. Comments from prominent officials suggest that the United States and the European Union may already be moving away from global multilateralism toward cooperation with limited circles of friends. In Janet Yellen's (2022) words:

> [W]e need to modernize the multilateral approach we have used to build trade integration. Our objective should be to achieve free but secure trade. We cannot allow countries to use their market position in key raw materials, technologies, or products to have the power to disrupt our economy or exercise unwanted geopolitical leverage. So let's build on and deepen economic integration . . . And let's do it with the countries we know we can count on. Favoring the friend-shoring of supply chains to a large number of trusted countries . . . will lower the risks to our economy as well as to our trusted trade partners.

In similar spirit, Christine Lagarde (2022), President of the European Central Bank, remarked:

> Russia's unprovoked aggression has triggered a fundamental reassessment of economic relations and dependencies in our globalised economy . . . Today, rising geopolitical tensions mean our global economy is changing . . . [O]ne can already see the emergence of three distinct shifts in global trade. These are the shifts from dependence to diversification, from efficiency to security, and from globalisation to regionalisation.

Competitor states, such as Russia and China, as well as "friendly" but nonallied states, such as India, have also begun looking for ways to disentangle their

economies from the current, American-dominated international financial system. One example of this has been Russia's insistence that payments for its natural gas exports be made in rubles (Hetzner 2022). Another example is China's efforts to establish the renminbi as a primary currency in cross-border transactions. If these countries transition to conducting business without having to rely on US institutions or the dollar, the economic power of the United States will diminish, and the effectiveness of its sanctions policies will be undermined.

Aggressive efforts to enhance the potency of sanctions—especially the expanded use of secondary trade sanctions and the US practice of asserting extraterritorial jurisdiction in sanctions—is also straining international relationships, even among longtime allies. European countries have expressed displeasure at the United States for enforcing its sanction and financial reporting laws on European firms, especially for business conducted with non-US enterprises.

While sanctions are more effective when imposed multilaterally, efforts to "coerce" allies to cooperate with sanctions policies can undermine the foundation of alliances. Ironically, efforts by the United States to leverage its economic strength in support of its sanctions policies may very well damage both its military and economic alliances—including the solidarity of the NATO alliance and the cohesion of the World Trade Organization—that have served as prominent anchors of its privileged position in world affairs. These relationships are enormously complex and interconnected; in contrast, the models at the core of our existing understanding of sanctions are typically assumed to view sanctions as dyadic, with one sender and one target. Yet even when we recognize that the sender may be a coalition of states, we generally assume, at least implicitly, that coalitional dynamics are exogenous to, and worked out before, bargaining occurs with the target (for a notable exception, see Miers and Morgan 2002). Further developments in our understanding of sanctions are likely to require the development of models that can account for the complexities in the formation of international coalitions and multilateral bargaining.[13]

After decades of experience with sanctions, both policymakers and researchers are still grappling with the basic question of whether sanctions are an effective tool of foreign policy. This makes the questions we have raised here all the more daunting. To understand sanctions processes more fully, we have to recognize their complexity. We also have to work to improve our understanding of the connections between economics and politics, the intricacies of multilateral bargaining, the degree to which instruments of policy (including the use of military force and sanctions) are substitutable or complementary, how behaviors and outcomes in one sanction incidence can affect expectations in future cases, and the dynamics of sanctions as cases unfold. Furthering our understanding of sanctions will be

[13] Eaton and Engers (1992) and Maggi (2016) contain numerous valuable ideas on the relationship between sanctions, issue linkage and bargaining. See also Anbarci, Skaperdas, and Syropoulos (2002) for an effort to capture the value of arming and bargaining protocols in negotiations conducted in the shadow of conflict.

challenging, but the recent sanctions on Russia have made it abundantly clear why we should make the effort.

■ *We wish to thank Eric Hurst, Nina Pavcnik, Timothy Taylor, and Heidi Williams for their outstanding suggestions and exemplary editorial guidance. We also wish to thank Navin Bapat, Michelle R. Garfinkel, Gladys J. Zubiria Fuentes, Aleksandra Kirilakha, Yoshiharu Kobayashi, Stergios Skaperdas, and Aiganym Valikhanova for their excellent comments.*

References

Allen, Susan Hannah. 2008. "Political Institutions and Constrained Response to Economic Sanctions." *Foreign Policy Analysis* 4 (3): 255–74.

Anbarci, Nejat, Stergios Skaperdas, and Constantinos Syropoulos. 2002. "Comparing Bargaining Solutions in the Shadow of Conflict: How Norms against Threats Can Have Real Effects." *Journal of Economic Theory* 106 (1): 1–16.

Attia, Hana, Julia Grauvogel, and Christian von Soest. 2020. "The Termination of International Sanctions: Explaining Target Compliance and Sender Capitulation." *European Economic Review* 129: 103565.

Bapat, Navin A., Tobias Heinrich, Yoshiharu Kobayashi, and T. Clifton Morgan. 2013. "Determinants of Sanctions Effectiveness: Sensitivity Analysis Using New Data." *International Interactions* 39 (1): 79–98.

Bapat, Navin A., and Bo Ram Kwon. 2015. "When Are Sanctions Effective? A Bargaining and Enforcement Framework." *International Organization* 69 (1): 131–62.

Bapat, Navin A., and T. Clifton Morgan. 2009. "Multilateral versus Unilateral Sanctions Reconsidered: A Test Using New Data." *International Studies Quarterly* 53 (4): 1075–94.

Barber, James. 1979. "Economic Sanctions as a Policy Instrument." *International Affairs* 55 (3): 367–84.

Barnett, Michael N. 1997. "Bringing in the New World Order: Liberalism, Legitimacy, and the United Nations." *World Politics* 49 (4): 526–51.

Besedeš, Tibor, Stefan Goldbach, and Volker Nitsch. 2017. "You're Banned! The Effect of Sanctions on German Cross-Border Financial Flows." *Economic Policy* 32 (90): 263–318.

Besedeš, Tibor, Stefan Goldbach, and Volker Nitsch. 2021. "Cheap Talk? Financial Sanctions and Nonfinancial Firms." *European Economic Review* 134: 103688.

Biersteker, Thomas, and Peter A.G. van Bergeijk. 2015. "How and When Do Sanctions Work? The Evidence." https://repub.eur.nl/pub/79414/On-target-Biersteker_Bergeijk-pp17-28.pdf.

Blum, Samantha. 2003. "Chinese Views of US Hegemony." *Journal of Contemporary China* 12 (35): 239–64.

Boudreau, Donald G. 1997. "Economic Sanctions and Military Force in the Twenty-First Century." *European Security* 6 (2): 28–46.

Camacho, Oscar R., Michelle R. Garfinkel, Constantinos Syropoulos, and Yoto V. Yotov. 2022. "Output Insecurity and Ownership Disputes as Barriers to Technology Diffusion." Drexel University, School of Economics Working Paper 2022-10.

Charnovitz, Steve. 2001. "Rethinking WTO Trade Sanctions." *American Journal of International Law* 95 (4): 792–832.

Choi, Seung-Whan, and Shali Luo. 2013. "Economic Sanctions, Poverty, and International Terrorism: An Empirical Analysis." *International Interactions* 39 (2): 217–45.

Cortright, David and George A. Lopez. 2000. *The Sanctions Decade: Assessing UN Strategies in the 1990s.*

Boulder, CO: Lynne Rienner.

Cortright, David and George A. Lopez. 2002. *Smart Sanctions: Targeting Economic Statecraft*. Lanham, MD: Rowman and Littlefield.

Demena, Binyam A., Alemayehu S. Reta, Gabriela Benalcazar Jativa, Patrick B. Kimararungu, and Peter A. G. van Bergeijk. 2021. "Publication Bias of Economic Sanctions Research: A Meta-Analysis of the Impact of Trade Linkage, Duration and Prior Relations on Sanctions Success." In *Research Handbook on Economic Sanctions*, edited by Peter A. G. van Bergeijk, 125–50. Cheltenham, UK: Edward Elgar Publishing.

Doxey, Margaret. 1972. "International Sanctions: A Framework for Analysis with Special Reference to the UN and Southern Africa." *International Organization* 26 (3): 527–50.

Drezner, Daniel W. 2000. "Bargaining, Enforcement, and Multilateral Sanctions: When Is Cooperation Counterproductive?" *International Organization* 54 (1): 73–102.

Early, Bryan R. 2011. "Unmasking the Black Knights: Sanctions Busters and Their Effects on the Success of Economic Sanctions." *Foreign Policy Analysis* 7 (4): 381–402.

Early, Bryan R. 2021. "Making Sanctions Work: Promoting Compliance, Punishing Violations, and Discouraging Sanctions Busting." In *Research Handbook on Economic Sanctions*, edited by Peter A. G. van Bergeijk, 167–86. Cheltenham, UK: Edward Elgar Publishing.

Eaton, Jonathan, and Maxim Engers. 1992. "Sanctions." *Journal of Political Economy* 100 (5): 899–928.

Eichenberg, Richard C. 2005. "Victory Has Many Friends: U.S. Public Opinion and the Use of Military Force, 1981–2005." *International Security* 30 (1): 140–77.

Elliott, Kimberly A. and Gary C. Hufbauer. 1999. "Same Song, Same Refrain? Economic Sanctions in the 1990's." *American Economic Review* 89 (2): 403–08.

European Commission. 2021. "The European Economic and Financial System: Fostering Openness, Strength and Resilience." https://ec.europa.eu/info/publications/210119-economic-financial-system-communication_en.

Farmer, Richard D. 2000. "Costs of Economic Sanctions to the Sender." *World Economy* 23 (1): 93–117.

Felbermayr, Gabriel, Aleksandra Kirilakha, Constantinos Syropoulos, Erdal Yalcin, and Yoto V. Yotov. 2020. "The Global Sanctions Data Base." *European Economic Review* 129: 103561.

Felbermayr, Gabriel, T. Clifton Morgan, Constantinos Syropoulos, and Yoto V. Yotov. 2021. "Understanding Economic Sanctions: Interdisciplinary Perspectives on Theory and Evidence." *European Economic Review* 135: 103720.

Felbermayr, Gabriel, Constantinos Syropoulos, Erdal Yalcin, and Yoto V. Yotov. 2022. "On the Heterogeneous Effects of Sanctions on Trade."

Galtung, Johan. 1967. "On the Effects of International Economic Sanctions: With Examples from the Case of Rhodesia." *World Politics* 19 (3): 378–416.

Garfinkel, Michelle R., Stergios Skaperdas, and Constantinos Syropoulos. 2015. "Trade and Insecure Resources." *Journal of International Economics* 95 (1): 98–114.

Garfinkel, Michelle R., Constantinos Syropoulos, and Yoto V. Yotov. 2020. "Arming in the Global Economy: The Importance of Trade with Enemies and Friends." *Journal of International Economics* 123: 103295.

Garfinkel, Michelle R., Constantinos Syropoulos and Thomas Zylkin. 2022. "Prudence versus Predation and the Gains from Trade." *Journal of Economic Theory* 201: 105434.

Gordon, Joy. 2016. "Extraterritoriality: Issues of Overbreadth and the Chilling Effect of Sanctions in the Cases of Cuba and Iran." *Harvard International Law Journal Online* 57 (1): 1–12.

Grossman, Herschel I., and Minseong Kim. 1995. "Swords or Plowshares? A Theory of the Security of Claims to Property." *Journal of Political Economy* 103 (6): 1275–88.

Gutmann, Jerg, Matthias Neuenkirch, and Florian Neumeier. 2020. "Precision-Guided or Blunt? The Effects of US Economic Sanctions on Human Rights." *Public Choice* 185 (1-2): 161–82.

Haidar, Jamal I. 2017. "Sanctions and Export Deflection: Evidence from Iran." *Economic Policy* 32 (90): 319–55.

Han, Baran. 2021. "Secondary Sanctions Mechanism Revisited: The Case of US Sanctions against North Korea." In *Research Handbook on Economic Sanctions*, edited by Peter A.G. van Bergeijk, 223–37. Cheltenham, UK: Edward Elgar Publishing.

Hetzner, Christiaan. 2022. "Why Putin Wants Russia to be Paid in Rubles." *Fortune*, April 1. https://fortune.com/2022/04/01/why-putin-wants-russia-to-be-paid-in-roubles/.

Hirshleifer, Jack. 1995. "Anarchy and Its Breakdown." *Journal of Political Economy* 103 (1): 26–52.

Hirschman, Albert O. 1945. *National Power and the Structure of Foreign Trade*. Berkeley and Los Angeles:

University of California Press.
Hoffmann, Fredrik. 1967. "The Functions of Economic Sanctions: A Comparative Analysis." *Journal of Peace Research* 4 (2): 140–59.
Hosoya, Chihiro. 1968. "Miscalculations in Deterrent Policy: Japanese-U.S. Relations, 1938–1941." *Journal of Peace Research* 5 (2): 97–115.
Hufbauer, Gary C. 1998. "Economic Sanctions." In *Proceedings of the ASIL Annual Meeting*, edited by John L. Hargrove, 332–35. Washington, DC: American Society for International Law.
Hufbauer, Gary C., and Euijun Jung. 2020. "What's New in Economic Sanctions?" *European Economic Review* 130: 103572.
Hufbauer, Gary C., and Thomas Moll. 2007. "Using Sanctions to Fight Terrorism." In *Terrornomics*, edited by Sean S. Costigan and David Gold, 179–94. London: Routledge.
Hufbauer, Gary C., and Jeffrey J. Schott. 1985. *Economic Sanctions Reconsidered: History and Current Policy*. Washington, DC: Peterson Institute.
Hufbauer, Gary C., Jeffrey J. Schott, and Kimberly A. Elliott. 1990. *Economic Sanctions Reconsidered: History and Current Policy*. 2nd ed. Washington, DC: Peterson Institute.
Hufbauer, Gary C., Jeffrey J. Schott, Kimberly A. Elliott, and Barbara Oegg. 2007. *Economic Sanctions Reconsidered: History and Current Policy*. 3rd ed. Washington, DC: Institute for International Economics.
Jacobson, Harold K., and Michel Oksenberg. 1990. *China's Participation in the IMF, the World Bank, and GATT: Toward a Global Economic Order*. Ann Arbor, MI: University of Michigan Press.
Jentleson, Bruce W., and Rebecca L. Britton. 1998. "Still Pretty Prudent: Post-Cold War American Public Opinion on the Use of Military Force." *Journal of Conflict Resolution* 42 (4): 395–417.
Kaempfer, William H. and Anton D. Lowenberg. 2007. "The Political Economy of Economic Sanctions." In *Handbook of Defense Economics*, Vol. 2, edited by Todd Sandler and Keith Hartley, 867–911. Amsterdam: Elsevier.
Kagan, Donald. 1969. *The Outbreak of the Peloponnesian War*. Ithaca, NY: Cornell University Press.
Kirilakha, Aleksandra, Gabriel J. Felbermayr, Constantinos Syropoulos, Erdal Yalcin, and Yoto V. Yotov. 2021. "The Global Sanctions Data Base (GSDB): An Update That Includes the Years of the Trump Presidency." In *Research Handbook on Economic Sanctions*, edited by Peter A.G. van Bergeijk, 62–106. Cheltenham, UK: Edward Elgar Publishing.
Kwon, Ohyun, Constantinos Syropoulos, and Yoto V. Yotov. 2022a. "Do Sanctions Affect Growth?" Drexel University, School of Economics Working Paper Series 2022-6.
Kwon, Ohyun, Constantinos Syropoulos, and Yoto V. Yotov. 2022b. "The Extraterritorial Effects of Sanctions." Drexel University, School of Economics Working Paper Series 2022-3.
Lagarde, Christine. 2022. "A New Global Map: European Resilience in a Changing World." Speech, Petersen Institute, Washington, DC, April 22, 2022. https://www.ecb.europa.eu/press/key/date/2022/html/ecb.sp220422~c43af3db20.en.html.
Lektzian, David and Mark Souva. 2007. "An Institutional Theory of Sanctions Onset and Success." *Journal of Conflict Resolution* 51 (6): 848–71.
Lektzian, David J., and Christopher M. Sprecher. 2007. "Sanctions, Signals, and Militarized Conflict," *American Journal of Political Science* 51 (2): 415–31.
Lindsay, James M. 1986. "Trade Sanctions as Policy Instruments: A Re-examination." *International Studies Quarterly* 30 (2): 153–73.
MacDonald, Brian R. 1983. "The Megarian Decree." *Historia: Zeitschrift für Alte Geschichte* 32 (4): 385–410.
Maggi, Giovanni. 2016. "Issue Linkage." In *Handbook of Commercial Policy*, Vol. 1B, edited by Kyle Bagwell and Robert W. Staiger, 513–64. Amsterdam: Elsevier.
Martin, Lisa L. 1992. *Coercive Cooperation: Explaining Multilateral Economic Sanctions*. Princeton, NJ: Princeton University Press.
Mastanduno, Michael. 2019. "Liberal Hegemony, International Order, and US Foreign Policy: A Reconsideration." *British Journal of Politics and International Relations* 21 (1): 47–54.
McGillivray, Fiona, and Allan C. Stam. 2004. "Political Institutions, Coercive Diplomacy, and the Duration of Economic Sanctions." *Journal of Conflict Resolution* 48 (2): 154–72.
Meyer, Jeffrey A. 2009. "Second Thoughts on Secondary Sanctions." *University of Pennsylvania Journal of International Law* 30 (3): 905–68.
Miers, Anne, and T. Clifton Morgan. 2002. "Multilateral Sanctions and Foreign Policy Success: Can Too Many Cooks Spoil the Broth?" *International Interactions* 28 (2): 117–36.
Mitchell, Andrew D. 2017. "Sanctions and the World Trade Organization." In *Research Handbook on*

UN Sanctions and International Law, edited by Larissa van den Herik, 283–303. Cheltenham, UK: Edward Elgar Publishing.

Morgan, T. Clifton. 1995. "Clinton's Chinese Puzzle: Domestic Politics and the Effectiveness of Economic Sanctions," *Issues and Studies* 31 (8): 19–45.

Morgan, T. Clifton and Navin A. Bapat. 2003. "Imposing Sanctions: States, Firms and Economic Coercion." *International Studies Review* 5 (4): 65–79.

Morgan, T. Clifton and Valerie L. Schwebach. 1997. "Fools Suffer Gladly: The Use of Economic Sanctions in International Crises." *International Studies Quarterly* 41 (1): 27–50.

Morgan, T. Clifton, Navin A. Bapat, and Yoshiharu Kobayashi. 2014. "The Threat and Imposition of Sanctions: Updating the TIES Dataset." *Conflict Management and Peace Science* 31 (5): 541–58.

Morgan, T. Clifton, Navin A. Bapat, and Yoshiharu Kobayashi. 2021. "The Threat and Imposition of Economic Sanctions Data Project: A Retrospective." In *Research Handbook on Economic Sanctions*, edited by Peter A.G. van Bergeijk, 44–61. Chetenham, UK: Edward Elgar Publishing.

Morgan, T. Clifton, Navin A. Bapat, and Valentin Krustev. 2009. "The Threat and Imposition of Economic Sanctions, 1971-2000." *Conflict Management and Peace Science* 26 (1): 92–110.

Morgan, T. Clifton, Constantinos Syropoulos, and Yoto V. Yotov. 2022. "Economic Sanctions: Evolution, Consequences, and Challenges." Drexel University, School of Economics Working Paper Series 2022-12.

Morgan, T. Clifton, Constantinos Syropoulos, and Yoto V. Yotov. 2023. "Replication data for: Economic Sanctions: Evolution, Consequences, and Challenges." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E183068V1.

Pape, Robert A. 1997. "Why Economic Sanctions Do Not Work." *International Security* 22 (2) 90–136.

Peksen, Dursun. 2019. "When Do Imposed Economic Sanctions Work? A Critical Review of the Sanctions Effectiveness Literature." *Defence and Peace Economics* 30 (6): 635–47.

Powell, Robert. 1993. "Guns, Butter, and Anarchy." *American Political Science Review* 87 (1): 115–32.

Russett, Bruce M. 1967. *International Regions and the International System: A Study in Political Ecology*. Chicago: Rand McNally & Company.

Ryan, Missy, and Annabelle Timsit. 2022. "U.S. Wants Russian Military 'Weakened' from Ukraine Invasion, Austin Says." *Washington Post*, April 25. https://www.washingtonpost.com/world/2022/04/25/russia-weakened-lloyd-austin-ukraine-visit/.

Schwebach, Valerie L. 2000. "Sanctions as Signals: A Line in the Sand or a Lack of Resolve?" In *Sanctions as Economic Statecraft*, edited by Steve Chan and A. Cooper Drury, 187–211. International Political Economy Series. London: Palgrave Macmillan.

Skaperdas, Stergios. 1992. "Cooperation, Conflict, and Power in the Absence of Property Rights." *American Economic Review* 82 (4): 720–39.

Syropoulos, Constantinos, Gabriel J. Felbermayr, Aleksandra Kirikakha, Erdal Yalcin, and Yoto V. Yotov. 2022. "The Global Sanctions Data Base— Release 3: COVID-19, Russia, and Multilateral Sanctions." Drexel University, School of Economics Working Paper Series 2022-11. https://www.globalsanctionsdatabase.com/.

van Bergeijk, Peter A. G. 2021. *Research Handbook on Economic Sanctions*. Cheltenham, UK: Edward Elgar Publishing.

Wagner, R. Harrison. 1988. "Economic Interdependence, Bargaining Power, and Political Influence." *International Organization* 42 (3): 461–83.

Weber, Patrick M., and Gerald Schneider. 2020. "How Many Hands to Make Sanctions Work? Comparing EU and US Sanctioning Efforts." *European Economic Review* 130: 103595.

Yellen, Janet L. 2022. "Special Address on: The Future of the Global Economy and US Economic Leadership." Speech, Atlantic Council, Washington, DC, April 13, 2022.

Zanchetta, Barbara. 2016. "Between Cold War Imperatives and State-Sponsored Terrorism: The United States and 'Operation Condor'." *Studies in Conflict & Terrorism* 39 (12): 1084–102.

Journal of Economic Perspectives—Volume 37, Number 1—Winter 2023—Pages 31–52

Financial Sanctions, SWIFT, and the Architecture of the International Payment System

Marco Cipriani, Linda S. Goldberg, and Gabriele La Spada

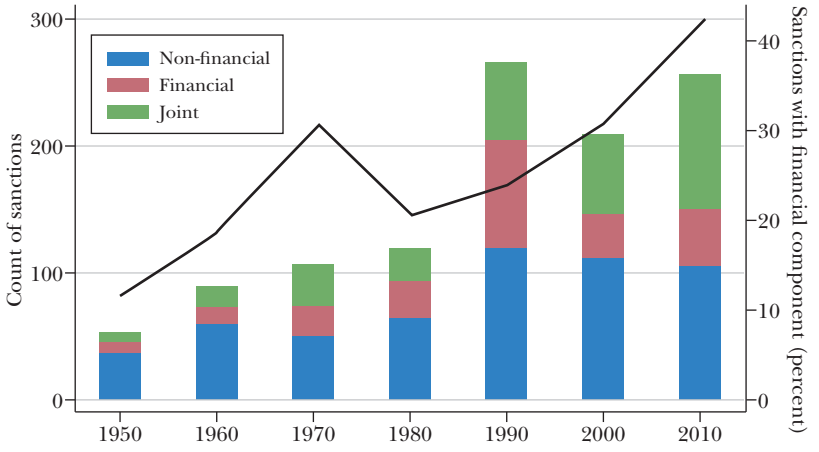
When sanctions involve traded goods, it is relatively easy to understand how they function. Either certain goods are permitted to cross a national border or they are not. In contrast, financial sanctions involve flows of funds, which occur through networks of banks and financial institutions. Financial sanctions typically restrict the ability of sanctioned entities—countries, businesses, or even individuals—to purchase or sell some financial assets. Sanctions can also be imposed on "custodial services," which refers to the ability of entities to store or manage the financial assets of the sanctioned entity. Other financial services, such as giving financial guidance or wealth management, can also be included.

Financial sanctions have been widely used for decades. Figure 1 shows the number of sanctions episodes by ten-year periods, from 1950 through 2019. Counts of sanctions have increased over time, from 52 sanctions episodes in the 1950s to 257 in the 2010s. The sanction type indicates whether an episode is characterized by the imposition of only economic, only financial, or jointly economic and financial sanctions. The share of financial sanctions has increased: the proportion of sanctions episodes with both a financial and a real economy component increased from of 12 percent in the 1950s to 42 percent in the 2010s. In contrast, exclusively economic sanctions decreased from

■ *Marco Cipriani is Head of Money and Payments Studies, Linda S. Goldberg is Financial Research Advisor, and Gabriele La Spada is a Financial Research Economist, all at the Federal Reserve Bank of New York, New York City, New York. Cipriani and Goldberg are both Research Fellows, Centre for Economic Policy Research, London, United Kingdom, and Goldberg is a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are marco.cipriani@ny.frb.org, linda.goldberg@ny.frb.org, and gabriele.laspada@ny.frb.org.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.1.31.

Figure 1
Prevalence of Sanctions By Type



Source: Authors' calculations using Global Sanctions Database (Felbermayr et al. 2020).
Notes: This figure displays the number of sanctions within each decade by sanction type. Sanctions are only counted in the first decade of implementation. Joint sanctions have both economic (for example, import restrictions) and financial (for example, asset freezes, investment screens) elements. The black line represents the percentage of sanctions with a financial component over time (right axis).

73 percent of the total in the 1950s to 41 percent in the 2010s. Exclusively financial sanctions were most prevalent in the 1980s and 1990s, reaching 32 percent of the total in the 1990s. Most sanctions are imposed by North American and European countries targeting Asian and African countries. Financial sanctions are more likely to be used than other sanctions when the goals are promoting democracy and human rights. On average, both financial and nonfinancial sanctions are imposed for shorter time periods now than in the past (Felbermayr et al. 2020).

In the last few decades, a particular type of financial sanction has become more prominent: restricting access to the infrastructures and institutions that execute international payments. This type of financial sanction can potentially disrupt every kind of cross-border economic activity requiring access to the payment system, including tourism, remittances, foreign exchange trading, and international trade financing. The vast majority of communications necessary for international payments is carried over the network maintained by the Society for Worldwide Interbank Financial Telecommunication (SWIFT), which has allowed for seamless flow of standardized information. Because the SWIFT system has very few alternatives, financial sanctions that limit access to this network have become particularly costly for sanctioned entities.

This article focuses on financial sanctions, with a particular emphasis on their relationship with the infrastructure of cross-border payments. We start with some background on financial sanctions since World War II, providing a number of

specific examples of sanctions episodes along with the historical context for their imposition and the types of activities included. We then describe the infrastructure of cross-border payments, before turning to the role of SWIFT in international financial markets and the use of financial sanctions restricting access to SWIFT. We conclude by discussing some of the alternative systems some countries have created to limit the dependency on this single network. While some of these alternative systems have achieved traction within their domestic economies, they have not yet gained broad use in cross-border activity.

Examples of Financial Sanctions since World War II

Several readily accessible sources provide details on sanctioned entities—for example, governments, businesses, or individuals—and the specific activities that are forbidden, how the sanctions are implemented, and which entities are tasked with administering sanctions. Across countries, the Global Sanctions Database produced by Felbermayr et al. (2020) and Kirilakha et al. (2021) divides sanctions into economic and financial sanctions, details which countries imposed the sanctions and the sanctioned entities, categorizes the purposes of the sanctions, reports their duration, and offers an assessment of sanction effectiveness.

For the United States, the Department of the Treasury provides specifics for financial sanctions on the Office of Foreign Assets Control (OFAC) website.[1] As one example, the OFAC defines "blocking a transaction" as: "blocking a transaction involves accepting or segregating the funds or securities associated with the transaction and then freezing those funds, securities or accounts so that the owner is effectively denied access until appropriate action is taken by OFAC. Blocking can occur when a transaction is initiated at an institution or when funds or securities are moved through an institution during a transfer."

The remainder of this section describes specific episodes of financial sanctions, explaining how and why sanctions were imposed, with the purpose of outlining the different forms that financial sanctions can take and how they have evolved over time.

US Sanctions against North Korea in the 1950s

In June 1950, North Korea invaded South Korea; in response to this attack, on June 25th and 27th, the United Nations Security Council passed Resolutions 82 and 83, sponsored by the United States, calling for North Korean authorities to withdraw and recommending urgent military measures by UN members. The United States imposed sanctions against North Korea in the 1950s, with the purpose of helping the United States win the Korean War.

[1] For details, see the Office of Foreign Assets Control website at https://home.treasury.gov/policy-issues/office-of-foreign-assets-control-sanctions-programs-and-information.

The US sanctions against North Korea had both trade and financial components. The trade restrictions, such as a total embargo on exports to North Korea, were instituted just three days after the outbreak of the war (on June 28, 1950). In addition to the embargo, the Department of the Treasury issued the Foreign Assets Control Regulations in January 1951, forbidding any financial transactions involving North Korea and its nationals. Moreover, the Department froze North Korean assets held under US jurisdiction (Chang 2006). China, an ally of North Korea in the war, was also subject to the same sanctions.

The United States did not impose restrictions on North Korean's access to the infrastructure allowing for international payments and financial transactions; rather, the US rules made such transactions illegal for US residents. Financial sanctions against North Korea are an example of sanctions where financial transactions and ownership of financial assets are impaired, but access of the country to the infrastructure of payment systems is not affected.

US Sanctions against Chile from 1970 to 1973

In the 1960s, Chile received extensive credit from the United States and from international organizations based in the United States, such as the Inter-American Development Bank. Indeed, in 1970, 60 percent of Chile's debt was owed to the US government (Helwege 1989). Moreover, in the late 1960s, private credit from the United States had become increasingly important, with US commercial banks providing significant lines of credits.

In 1970, Salvador Allende won the Chilean presidential election and started pursuing domestic and international policies contrary to US interests in the region. Between 1970 and 1973, the United States put in place a series of economic measures against Chile which became known as "the invisible blockade," aimed at destabilizing the country and overthrowing Allende (Petras and Morley 1975; 1978; Olson 1979).

In addition to trade restrictions, financial activity between the United States and Chile decreased significantly. The United States tightened official-sector credit flows towards Chile: US Agency for International Development loans were reduced from $45 million in 1969 to $1.5 million in 1971, and Import-Export Bank credits evaporated entirely. In addition, the Inter-American Development Bank reduced the credit provided to Chile from $46 million in 1970 to $2 million in 1972 (Livingstone 2009).

US private financing also declined dramatically: short-term lines of credit from US private banks declined to around $30 million, and short-term US commercial credits dropped from 78.4 percent of the total in 1970 to approximately 6.6 percent in 1972 (Petras and Morley 1975; 1978). Additionally, US suppliers were demanding "cash in advance" for essential raw materials and parts sales to Chile, putting further pressure on Chile's finances (Petras and Morley 1975; Olson 1979; Livingstone 2009). The reduction in private-sector lending may have been due both to the nature of the policies put forward by the Allende government—which were generally not business-friendly (for example, completing the nationalization of American

copper companies in Chile that began in 1965)—and to the desire of US financial institutions to be aligned with the policies of the US government (Sigmund 1974; Petras and Morley 1978).

As in the case of the sanctions against North Korea, the United States did not target the infrastructure of financial transactions. However, different from the North Korean case, the United States neither adopted explicit measures forbidding financial transactions between the United States and Chile nor froze the US financial assets owned by Chilean residents. Instead, the US government relied on the economic disruption brought forward by a little-publicized reduction in both official-sector and private-sector lending to the country (Olson 1979). The US economic pressure on Chile ended after the military coup that overthrew the Allende's government in late 1973.

European and US Sanctions against South Africa in the 1980s and 1990s

Since the early 1960s, the United Nations and many countries called for and implemented economic sanctions against South Africa in order to pressure the South African government to abandon its apartheid policy of racial segregation (Crawford and Klotz 1999). In 1963, the UN Security Council adopted a voluntary arms embargo, which it made mandatory in 1977. In November 1973, the OPEC counties extended their oil embargo to South Africa.

Although some financial sanctions were put in place in the 1960s and 1970s (for example, Japan banned direct investment in 1964 and then loans in 1975), more extensive financial sanctions were introduced during the South African debt crisis of 1984–1985, along with a tightening of the trade-based economic sanctions. In 1986, the European Community, the United States, and Japan sanctioned import of gold coins (the South African Krugerrand) and certain steel and iron products. However, most forms of gold, which accounted for 42.6 percent of the value of South African merchandise exports, were not sanctioned (Crawford and Klotz 1999; Levy 1999). Financial sanctions mainly focused on foreign direct and portfolio investments in South Africa. The European Community sanctioned new direct investments, but member states were not required to impose binding sanctions; indeed, Great Britain and Germany—the two major investors in South Africa—decided not to do so (Crawford and Klotz 1999; Becker 1988; Hefti and Staehelin-Witt 2011). The United States sanctioned new direct investments through the Comprehensive Anti-Apartheid Act (CAAA) of 1986, along with portfolio investments and credits and loans. The CAAA also prohibited US banks from accepting deposits from South African government agencies (Becker 1988).

The South African apartheid regime ended with the general election of 1994. Similarly to the United States's sanctions against North Korea, the 1980s financial sanctions against South Africa did not involve the infrastructure of cross-border payments; indeed, they were more limited than sanctions on North Korea, targeting mainly foreign direct investment into South Africa. Moreover, similar to other cases described above, sanctions were accompanied by significant actions by nongovernment actors, such as divestment by US universities and pension funds from companies

doing business in South Africa. As a result of both pressure from the anti-apartheid movement and the concerning conditions of South African economy, several banks and multinational companies disinvested from South Africa in the 1980s. For example, in July 1985, Chase Manhattan Bank decided not to extend credit or to make new loans to South Africa; immediately after, other international banks and investors moved their funds out of the country, leading the Johannesburg Stock Exchange to drop sharply and the South African rand to plummet. In 1986, in response to customer pressures, Barclays Bank ended its loans to South Africa and withdrew from South African operations. In the same year, General Motors withdrew from South Africa, followed by many other US corporations (Crawford and Klotz 1999).

Although the sanctions' goal was ultimately achieved, the contribution of foreign economic and financial pressures to the regime's downfall is still debated (Levy 1999). During the sanctions period (1986Q4 to 1991Q1), South Africa suffered an average net capital outflow of 2 percent of South African GNP. However, this outflow seems mostly attributed to poor economic conditions rather than to the impact of sanctions. Although sanctions made capital scarce, the annual cost to the South African economy is estimated at less than 0.25 percent of South African GNP. The relatively low effectiveness is attributed to the lack of sanctions by the UK and Germany, to the fact that sanctions did not cover reinvested profits (80 percent of FDI into South Africa), and to the fact that only the United States sanctioned portfolio investment (Hefti and Staehelin-Witt 2011).

EU and US Sanctions against Myanmar in the 1990s and 2000s

In the 1990s and 2000s, the European Union and the United States adopted several economic sanctions against Myanmar in response to systematic violations of human rights and civil liberties by the country's ruling military junta. In 1991, the European Union imposed an array of traditional economic sanctions, including an arms embargo, a suspension of bilateral aid, and a visa ban on Myanmar officials (Giumelli and Ivan 2013). In 2000, the European Union strengthened the existing economic sanctions and added a financial component, freezing the funds held abroad by the persons included in the visa ban. In 2004, the European Union imposed restrictions on EU investment into Myanmar, in particular into Burmese state-owned firms (European Commission 2006). Similar restrictions on investment into Myanmar were introduced by the United States. Finally, Canada, the European Union, and the United States stopped providing preferential financing for exports to or investment in the country (Martin 2012).

Myanmar is a significant example of international pressure to impose restrictions on a country's access to the infrastructure of the financial system. Beginning in 2004, human rights groups like Human Rights Watch urged SWIFT to remove Myanmar banks owned by the ruling military junta from its network, pointing out that the military dictatorship could use the network to evade the economic and financial sanctions. In this instance, SWIFT refused to disconnect the banks, in order to maintain an apolitical posture, on the ground that no EU law restricted access to SWIFT by Myanmar (Wong and Nelson 2021).

US Sanctions against Afghanistan in the 2000s

In the late 1990s and early 2000s, several countries and international organizations imposed economic and financial sanctions against the Taliban regime ruling Afghanistan. The goal of these sanctions was to force the Afghan government to stop sheltering and training terrorists. These sanctions were aimed at putting pressure not only on segments of the economy but also on specific individuals. An example of financial sanctions imposed against Afghanistan is US Executive Order No. 13129, issued in July 1999, banning all trade with Taliban-controlled areas, freezing Taliban assets in the United States, and prohibiting financial contributions to the Taliban (Hufbauer et al. 2001).

Shortly afterwards, in October 1999 and December 2000, the UN Security Council adopted two rounds of sanctions against the Taliban regime (Council Resolutions 1267 and 1333; see Francioni and Lenzerini 2003; Ghufran 2001). The Council's actions included travel bans, an arms embargo, and a ban on exports of acetic anhydride, used to manufacture heroin (of which Afghanistan was the world's largest producer). Finally and most importantly, these sanctions froze funds and other financial assets, owned directly or indirectly, by the Taliban, Osama bin Laden, and individuals transacting with him. One of the main goals of these sanctions was to coerce the Taliban to hand over Osama bin Laden. The effectiveness of these sanctions, however, is still debated, as the Taliban did not turn over bin Laden nor did al-Qaeda stop its terrorist activity.

Sanctions against Afghanistan intensified after the attacks of September 11, 2001. On September 23, 2001, in order to weaken the financial support of al-Qaeda, the US President George W. Bush issued an executive order expanding the list of individuals and entities subject to the asset freeze, including fundraising organizations (Hardister 2002). Reducing the financial capabilities of terrorist organizations was seen as a key component of the "war on terrorism." Moreover, the United States created the Foreign Terrorist Asset Tracking Center in the Treasury Department to coordinate the activities of the US agencies on the financial front.

Importantly, the US "war on terrorism" included a covert monitoring of global financial transactions through the SWIFT network (Connorton 2007; Koppel 2011). In October 2001, the US Treasury established a secret program—later referred to as the "Terrorist Financing Tracking Program" (TFTP) but more commonly known as the "SWIFT Program'"—through which the Office of Foreign Assets Control (OFAC) would issue subpoenas to the SWIFT data processing center in the United States. The amount and type of data accessed by US authorities is not publicly known. As SWIFT acknowledged, initially the scope of US searches covered the entire SWIFT database, with the transfer to the US Treasury of all messages within a certain time period. Subsequent subpoenas, however, were narrower and limited to specific dates and countries of origin or destination (Koppel 2011). According to US Treasury officials, in 2007–2008, US counterterrorism analysts at the Central Intelligence Agency (in charge of extracting individual-level information from the SWIFT messages) have searched less than 1 percent of the subset of SWIFT messages sent to the US Treasury (Amicelle 2011).

The existence of the program became public in 2006, following a series of articles in major US newspapers. Although the program was legal under US law, it generated controversies both in the United States and European Union because of its implications for privacy and civil liberties (De Goede 2012). In particular, the Treasury received details about millions of messages, including sender's and receiver's personal data. Although the data were obtained from the US-based SWIFT center, they contained information on non-US citizens too, and European authorities expressed serious concerns about possible violations of European privacy law (Amicelle 2011; Koppel 2011; De Goede 2012). Negotiations between the United States and European Union, combined with mounting media pressure, led to an agreement on the SWIFT surveillance program between the United States and the European Union in June 2007, limiting the use of the data by US authorities to counterterrorism purposes, limiting the retention period for the data to five years, and allowing monitoring of the program by EU officials (Connorton 2007; Koppel 2011).

Moreover, EU pressures led SWIFT to improve its data protection standards and to create two message-processing zones: one in Europe (with processing centers located in the Netherlands and Switzerland) and one in North America (with processing centers located in the Netherlands and the United States), thereby separating EU traffic and US traffic. Countries have the option to choose which processing zone (and therefore pair of processing centers) they want to belong to. This change means that all traffic within the European processing zone, to which most countries have opted to belong, is not accessible by US surveillance.

The Infrastructure of Cross-Border Payments

Cross-border payments infrastructures are a critical component of how governments, companies, and households actually can pay for their international purchases, whether of goods, services, or financial assets. Once we understand this infrastructure, it becomes clear why restricting access to the infrastructure has been part of several sanction packages, especially in the most recent years, including the 2022 sanctions against Russia discussed in more detail later in this paper.

Payments within a single country typically settle on the books of commercial banks or of the central bank. For instance, if entity X wants to send funds to entity Y, X will instruct X's own bank, which will make a payment to Y's bank. If both X and Y are customers of the same bank, the payment can be settled on the bank's own books. If, however, X and Y are customers of different banks in the same country, the settlement will typically occur on the books of the central bank. Many central banks, in fact, have set up "real time gross settlement (RTGS)" systems that allow the settlement of payments between banks in real time, on a gross basis. In the United States, Fedwire Funds Services is a notable example. In the case of X and Y just described, the payment will result in a decrease of the account balances of X's bank with the central bank and an increase in the account balances of Y's bank.

Now consider cross-border payments, which are payments between residents located in different countries. Few central banks allow their domestic payment system to be accessed by banks that do not have a physical presence within the country and that are not subject to the country's regulation and supervision; one notable example is the Swiss National Bank, which allows access to its real-time gross settlement system (Swiss Interbank Clearing) to institutions without a presence in the country. Traditionally, however, cross-border payments occur through "correspondent banking": banks use the services of "correspondent banks" in order to execute cross-border payments (Bank for International Settlements 2016). The correspondent bank is usually either a large bank or a local branch/subsidiary of the bank initiating the payment, located in the foreign country where the payment must be sent. Banks may have more than one correspondent bank in a given country.[2]

A "correspondent account" is an account that a "respondent bank" has at a foreign correspondent bank usually in the foreign bank's currency. Both banks will keep a record of this account, and common terminology here is to refer to *Nostro* and *Vostro* accounts, where the terms are the Italian words for ours and yours. The record kept by the respondent bank of the money that it keeps with its correspondent bank is the *Nostro* account, whereas the record kept by the correspondent bank of its respondent bank's money is the *Vostro* account. Panel A of Figure 2 illustrates how a correspondent-bank relationship would allow a payment between parties X and Y in different countries. In this example, if X wants to send money to Y, who resides in another country, X's bank will instruct its correspondent bank in the country where Y resides to send money to Y's bank through the domestic payment system of Y's country. The *Nostro* account of X's bank (an asset on the bank's balance-sheet) will be debited for the amount paid; similarly, the *Vostro* account (a liability on the correspondent bank's balance sheet) will also be debited.

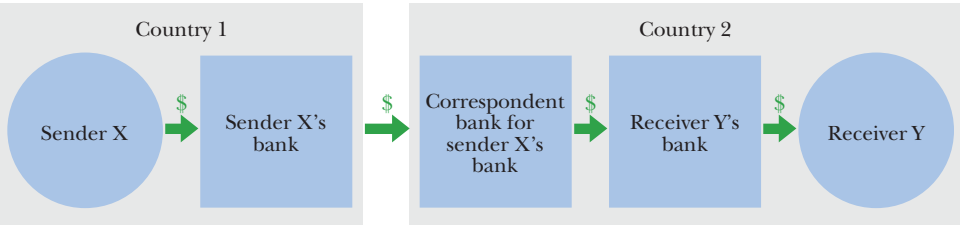
Cross-border payments through a correspondent banking relationship may be more complex than in the previous example and involve more than one intermediary. For instance, say that X and Y agree that the payment should be in a third country's currency. As shown in panel B of Figure 2, X's bank will instruct its correspondent bank in the third country to transfer funds (through the third country's domestic payment system) to the correspondent bank of Y's bank in the third country. In some cases, many correspondent banks are involved in the settlement of a payment. As correspondent banks are compensated for their correspondent services, the higher is the number of banks involved, the higher is the cost of the transaction.

Correspondent banks typically perform their own diligence for anti-money laundering (AML) and countering the financing of terrorism (CFT) purposes based on the requirements of their jurisdictions. The Financial Action Task Force (FATF), an intergovernmental organization established in 1989 by the G-7 countries (Canada, France, Germany, Italy, Japan, United Kingdom, and the United

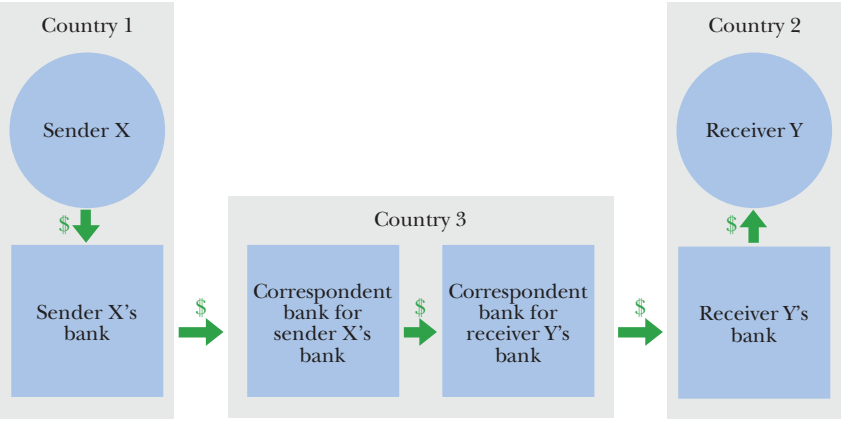
[2] Correspondent banking is also used for domestic payments when a bank does not have an account with the central bank; for instance, in the United States, community banks and credit unions often do not have an account with the Federal Reserve.

Figure 2
Cross-border Payments and Correspondent Banking

Panel A. Two countries



Panel B. Three countries



Source: Authors' construction.
Note: Panel A shows the flows in a cross-border payment executed through a correspondent bank. Panel B shows the flows in a cross-border payment executed through correspondent banks domiciled in a third country.

States), develops AML/CFT guidance on effective supervision and enforcement. Normally, policies with regard to AML/CFT require institutions to conduct due diligence on their respondent banks, but not on the customers of their respondent banks; nevertheless, correspondent institutions are usually required to monitor respondent banks' transactions "with a view to detecting any changes in the respondent institution's risk profile or implementation of risk mitigation measures" (FATF 2016). For instance, in the United States, banks are generally required to collect information on the origin and the recipient of transactions. Sometimes, correspondent banks may be held liable by the authorities of the country where they are located for violations of laws or regulations aimed at AML/CFT by their respondent banks.

Compliance with laws and regulations that are targeted at anti-money laundering or countering the financing of terrorism often makes correspondent banking relationships very costly for the correspondent bank. This is especially true if the respondent bank is located in a small country, where the volume of transactions is low, or in a country deemed at high risk for compliance with the rules that concern AML/CFT. The high cost and low profitability have resulted in a decrease in the number of correspondent banking relationships over the last decade, a phenomenon called "de-risking" (Grolleman and Jutrsa 2017; Miller 2022). The number of active correspondent banks worldwide fell by roughly 22 percent between 2011 and 2019, with banks losing correspondents even as the value of cross-border payments continued to grow; the decline has been especially pronounced in Latin America. At the same time, the number of country-pairs linked by a correspondent relationship—the so-called "corridors"—has decreased by 10 percent, leaving some regions, especially in Latin America, Oceania, and Africa, with very few corridors (Rice, von Peter, and Boar 2020; Bank for International Settlements 2020).

The reduction in correspondent banking relationships implies that some cross-border payment activity, especially if it involves small countries, needs to go through a longer chain of banks, potentially increasing the cost to end-user. Although comprehensive data on the cost of correspondent banking is lacking, the World Bank collects data on the cost of remittances from migrants, with the aim of reducing it to promote financial inclusion; although the cost of remittances from the high-income G-20 countries has been steadily decreasing since the 2010s, the concern is that de-risking by banks may slow down the process, or even lead to higher costs of remittances in some countries (World Bank 2022).

The high cost of correspondent banking activities has also led to the development of alternative arrangements to facilitate payment activity between residents of different countries (for a detailed analysis of these arrangements, see Bech and Hancock 2020). These arrangements may be sponsored by a single country or may be the result of multilateral agreements among a group of countries. For instance, a country may want to allow its residents to send and receive payments from a larger economy or currency area. One example is Switzerland, which established the Swiss Euro Clearing Bank (SECB) to allow its residents to send and receive euro payments from the European Union. Similarly, the Central Bank of Mexico, in a joint effort with the Federal Reserve, established "Directo a Mexico" to connect its own payment system to that of the United States.

Conversely, a country may want to set up a system to facilitate the use of its own currency by foreign residents or to facilitate regional transactions. As a prominent example, in 2015, China established the Chinese Cross-Border Interbank Payment System (CIPS) to facilitate the use of the renminbi in international transactions. As CIPS has developed a messaging system alternative to SWIFT, we will discuss it at more length in the next section, which describes the SWIFT messaging network and its origins, as well as its use in financial sanctions.

Sometimes a group of countries, usually neighbors, may jointly develop a payment system to allow their residents to transact among themselves; for example, the Southern Africa Development Community, a group of 16 countries in southern Africa, set up its own real-time gross settlement system for transactions in South African rand. Another example is the East Africa Payment System (EAPS), which offers multi-currency payments for countries in the East Africa Community, which includes Burundi, Kenya, Rwanda, Tanzania, and Uganda.

Even if both parties to a transaction are residents of the same country, there are cases in which payment in another country's currency may require the costly intermediation of correspondent banks. This friction has prompted some jurisdictions to set up "offshore" payment systems, processing payments in a currency different from that of the country where the payment system is based. For instance, Hong Kong has set up parallel real-time gross settlement systems that, in addition to the Hong Kong dollar, will also settle payments in euros, US dollars, and Chinese renminbi.

Finally, foreign-exchange transactions pose a particular type of settlement risk, as they require the payment of an agreed amount in one currency against an agreed amount in another currency.[3] What is called "payment versus payment" settlement mitigates settlement risk by only allowing two legs of a foreign-exchange transaction to settle contemporaneously. The CLS Bank, based in the United States, is a specialized financial intermediary established in 2002 to allow the settlement of foreign exchange transactions on a payment versus payment basis; it currently allows for foreign-exchange transactions in 18 currencies (Galati 2002).[4] CLS has 70 members, which are major financial institutions that hold accounts with CLS, and it settles the transactions between its members on its books.

A History of SWIFT

In the middle of the twentieth century, banks communicated nationally and internationally through Telex. For readers who have not yet reached retirement age, the Telex is a teleprinter network that originally used existing telegraph and telephone networks and allowed speech and teleprinter signals on the same connection. Introduced in the 1930s, the Telex quickly replaced the telegram in business use and grew fast in popularity: by 1957, there were more than 30,000 users worldwide, and in the late 1970s, more than one million. However, Telex messages were

[3] Settlement risk in foreign-exchange transaction is usually referred to as "Herstatt Risk": in 1974, Herstatt Bank, a German bank active in foreign-exchange trading, was closed by German authorities after it had received payments for foreign-exchange transactions, but before it could make the outgoing payments, leading to a freeze in the foreign-exchange market.

[4] The 18 currencies are: Australian dollar, Canadian dollar, Danish krone, euro, HK dollar, Hungarian forint, Israeli shekel, Japanese yen, Mexican peso, New Zealand dollar, Norwegian krone, Singapore dollar, South African rand, South Korean won, Swedish krona, Swiss franc, UK pound sterling, and US dollar.

costly and carried high operational risk: because Telex communication allowed to send unformatted texts with no prespecified standard, a cross-border transaction would often require the exchange of more than ten messages—and authentication procedures between banks were also labor intensive (Scott and Zachariadis 2012).

By the early 1970s, there was a growing presence of European and US banks in overseas markets and a rise in cross-border payment activity. Banks began looking for ways around the high costs and other limitations of Telex. In one prominent example, in 1973 Citibank's information technology subsidiary (Transaction Technology Inc.) developed a proprietary messaging system called MARTI (Machine Readable Telegraphic Input). By mid-1974, this network was in place and a pilot implementation had been conducted with one of Citibank's correspondent banks, Wilmington Trust. Citibank tried to force the adoption of MARTI on other correspondent banks, both in the United States and Europe, announcing that the deadline for compliance would be March 31, 1975. Many correspondent banks, particularly in Europe, resisted the imposition of a proprietary standard from a single bank (Scott and Zachariadis 2014). Indeed, European banks feared the establishment of a US-led monopoly for the transmission of financial information.

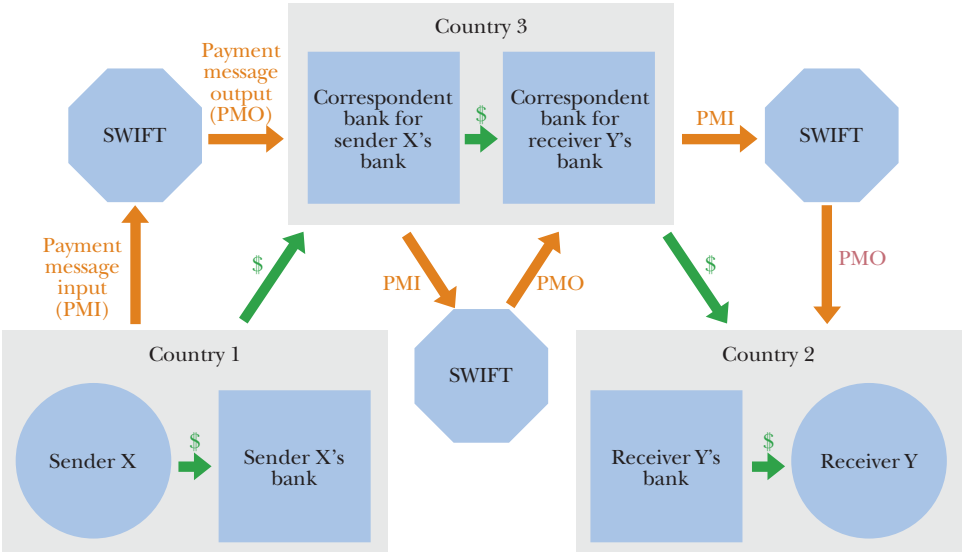
Thus, in 1973, 239 banks from 15 countries founded SWIFT, the Society for Worldwide Interbank Financial Telecommunication, as a nonprofit financial institution. The goal was to create a data processing and messaging network that would be shared among banks worldwide, with standards collectively designed by private companies for community purposes (Scott and Zachariadis 2014). SWIFT is headquartered in Belgium and is organized as a cooperative society owned by its members; membership was originally limited to banks but is now open to broker-dealers and investment management institutions.[5]

Figure 3 shows the role of SWIFT in the correspondent bank transaction we described earlier and illustrated in Figure 1. The primary role of SWIFT is as a message carrier: the SWIFT network securely transports messages containing the payment instructions between financial institutions involved in a transaction. In addition to providing the messaging network for financial transactions, SWIFT offers a secure person-to-person messaging service for the transfer of sensitive business documents, like contracts and invoices.

Importantly, SWIFT is not a bank and does not manage accounts or hold funds on behalf of its customers. It is also not a clearing or settlement institution. SWIFT only provides the platform allowing the secure exchange of financial information and proprietary data across financial institutions worldwide. Namely, SWIFT provides two main services to the financial sector: (1) a secure network for transmitting messages between financial institutions; and (2) the development and maintenance of a set of syntax standards for financial messages (Scott and Zachariadis 2012). For this reason, SWIFT does not eliminate the role of correspondent banks and other institutions involved in the settlement process.

[5]For details, see the SWIFT website at https://www.swift.com/about-us/legal/corporate-matters/swift-user-categories#shareholding-eligibility.

Figure 3
The Role of SWIFT in Cross-border Payments



Source: Authors' construction.
Note: This figure replicates Panel B of Figure 2 highlighting the role of the SWIFT network in facilitating cross-border payments.

SWIFT's messaging network is run from three data centers, located in the United States, the Netherlands, and Switzerland: as we mentioned above, these three centers create two separate message-processing zones in Europe and in North America. SWIFT uses undersea fiber-optic communications cables to transmit financial data across countries (Sechrist 2010). The CLS Bank, which as we described earlier operates the largest multi-currency cash settlement system, conducts millions of transactions and trades for trillions of US dollars a day on the same undersea cables. SWIFT's data centers share information in near real-time; in case of a failure in one of the data centers, the other centers are able to handle the traffic of the whole network.

The other fundamental purpose of SWIFT has been the development of a set of syntax standards that would facilitate financial transactions, overcoming the high processing costs and low reliability associated with Telex and easing information transmission. New standards are continuously developed and replace older ones. For example, ISO 9362, developed in 1994, defines a standard format for Business Identifier Codes (BIC) to uniquely identify financial and nonfinancial institutions worldwide; ISO 10383, developed in 2003, defines codes for exchanges and market identification; ISO 13616, developed in 2003, defines the International Bank Account Number (IBAN) to uniquely identify bank accounts worldwide; and

ISO 20022, developed by updating earlier standards in 2004 and 2007, defines a universal message scheme for electronic data interchanges between financial institutions, including payments, credit and debit card transactions, as well as securities trading and settlement. These formats are currently the main standards used in financial transactions.

SWIFT's standardization of financial messages has become the most influential and widely used in the financial industry. Indeed, the International Organization for Standardization (ISO) appointed SWIFT as the Registration Authority (that is, the entity responsible for defining and maintaining the rules) for several ISO standards. Currently, there are nine broad categories of SWIFT messages, ranging from funds transfers to foreign exchange transactions.[6]

Although SWIFT shareholders can only be banks, broker-dealers, and investment management institutions, the network can be used by a much broader set of institutions, including any supervised financial institution, international or intergovernmental bodies involved in finance and payments, nonsupervised financial institutions, corporations, financial market regulators, payment systems, and security market infrastructures.[7] Today, shareholders represent only roughly one-fourth of users.

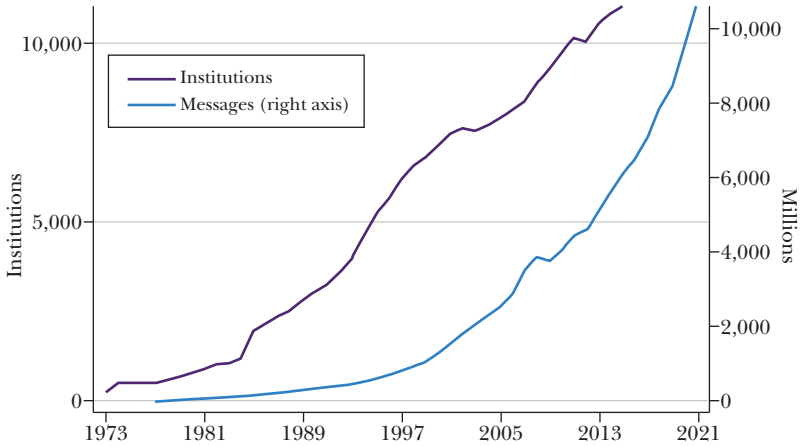
As Figure 4 shows, usage of the SWIFT network has grown steadily: in 2020, more than 11,000 institutions, located in more than 200 countries, were connected to SWIFT. In 2020, more than 9.5 billion messages were sent through the network, with an average daily volume of 37.7 million messages. Roughly 49 percent of this traffic was for securities trading and 45 percent for payments. The share of messages regarding securities trading has steadily been increasing over time, going from 40 percent in 2007 to almost 50 percent in 2020. As to the location in which traffic originated, 27 percent originated in the Americas, 59 percent in the region comprising Europe, the Middle East, and Africa, and the rest in the Asia-Pacific region (SWIFT 2020). At the end of 2020, 39 percent of total payment value was sent in US dollars and 37 percent in euros; in terms of trade finance, the US dollar represented 86 percent of the value of SWIFT traffic, the euro only 7 percent, and the Chinese renminbi amounting to 2 percent of the total value (SWIFT 2021).

Because SWIFT is not a payment or settlement system, the National Bank of Belgium does not regulate it as such. Since the late 1990s, however, it has been subject to the oversight of the Belgian central bank, together with the other central banks of the Group of Ten (G-10) countries and the European Central Bank, as a critical service provider. The oversight primarily focuses on the systemic risks related

[6]These nine categories are Customer payments and cheques (MT1XX); Financial institution transfers (MT2XX); Treasury markets—Foreign exchange and derivatives (MT3XX); Collection and cash letters (MT4XX); Securities Markets (MT5XX); Treasury markets—Precious metals and syndications (MT6XX); Documentary credits and guarantees (MT7XX). Traveller's cheques (MT8XX); and Cash management and customer status (MT9XX). Table A1 in the Appendix shows a fictitious example of transaction involving one of SWIFT's most common messages, the MT103 funds transfer message.

[7]For detailed user and shareholder eligibility criteria see the SWIFT website at https://www.swift.com/node/7776. The expansion of the eligibility criteria to use SWIFT started in the late 1980s.

Figure 4
Number of Institutions Connected to SWIFT and Number of SWIFT Messages



Source: Authors' calculations using data from Scott and Zachariadis (2014) and from SWIFT Annual Reviews from 2007 to 2021.
Note: The black line reports the number of institutions connected to SWIFT; the blue line reports the number of SWIFT messages (right axis).

to the confidentiality, integrity, and availability of the SWIFT network. In 2012, the SWIFT Oversight Forum, including an additional 15 central banks, was set up to increase information sharing on oversight activities. Because of the complexity of SWIFT regulatory regime, the impact of central banks' oversight mainly occurs through cooperation and moral suasion.

Over the years, SWIFT has completely displaced the systems that were previously used for communicating across financial institutions and across borders; indeed, most countries have discontinued their Telex communications services in the last decade or so. In a nutshell, SWIFT has become a critical institution for international payments without any major competitors.

The Role of SWIFT in the Implementation of Financial Sanctions

SWIFT has two key roles: allowing SWIFT participants to exchange information through the SWIFT network and setting standards for messaging. International sanctions involving SWIFT prevent sanctioned entities from accessing the SWIFT network. As we discuss below, because SWIFT standards are public, sanctions cannot prevent countries from developing parallel systems that employ SWIFT standards.

Furthermore, because SWIFT is a cooperative, its mission is to act in the interest of its entire member community. As such, SWIFT typically tries not to make policy decisions that exclude users or restrict their access to the platform. Decisions to

impose sanctions belong to the governments of countries, and governments around the world may (and do) impose very different sets of sanctions. As it is incorporated under Belgian law, SWIFT must comply with Belgian and EU laws and follow the sanction regimes under those jurisdictions.[8]

On some occasions, SWIFT has resisted political pressure to disconnect a country from its messaging system. For instance, in 2004, SWIFT resisted the call from human rights groups to remove Myanmar from its network even after the United States and the European Union had imposed sanctions on the country. In 2014, SWIFT resisted pressure from pro-Palestinian groups to disconnect Israeli financial institutions. With these actions, SWIFT reaffirmed its commitment to function as a neutral financial service provider.

However, in February 2012, the United States passed the Iran Sanctions, Accountability, and Human Rights Act of 2012, authorizing the US president to impose sanctions on persons or institutions that provided financial messaging services to designated Iranian financial institution, including SWIFT. As a response to the US legislation, SWIFT announced the decision to discontinue access to designated Iranian financial institutions as soon as it had clarity from the European Union. On March 15, 2012, the European Union passed EU Regulation 267/2012, forbidding SWIFT from providing financial messaging services to some EU-sanctioned Iranian banks, including Iran's central bank. SWIFT complied with this regulation and disconnected the EU-sanctioned Iranian banks from its system.

The imposition of a financial sanction via SWIFT can also happen without any legislative action from Belgium or the European Union. In 2017, Belgium decided it would no longer allow SWIFT to provide services to certain UN-sanctioned North Korean banks, and SWIFT removed these institutions. The following week, SWIFT disconnected the remaining North Korean banks, without being required to do so by either Belgian or EU law. Although SWIFT offered an explanation for this follow-up decision by saying that the remaining banks had failed to meet its operating criteria, it did not explain what exactly the banks did that justified the suspension.

Similarly, SWIFT may decide to follow the directives of a country even if it is not required to do so by Belgian and EU law. In 2016, Iranian banks were reconnected to SWIFT following the Joint Comprehensive Plan of Action—also known as the Iran nuclear deal of 2015—agreed to by Iran, China, the EU, France, Germany, Russia, the United Kingdom, and the United States. When the United States withdrew from the deal in 2018, it gave SWIFT a six-month period to disconnect the re-sanctioned Iranian institutions, or face US sanctions. Because the European Union had not withdrawn from the treaty, EU regulation did not force SWIFT to disconnect the Iranian financial institutions re-sanctioned by the United States. After the six-month period ended, however, SWIFT decided to disconnect the Iranian banks from its system "in the interest of the stability and integrity of the wider global financial

[8]This is explained at the SWIFT website at https://www.swift.com/about-us/legal/compliance-0/swift-and-sanctions.

system." US sanctions on SWIFT would have imposed a significant impact on the global economy, given the centrality of SWIFT to the global payment system.

In 2014, following Russia's annexation of the Crimea region of Ukraine, the United States, the European Union, and Canada introduced targeted sanctions against Russian individuals and entities, mainly travel restrictions, asset freezes, and restrictions on debt and equity financing. On September 18, 2014, the European Parliament also passed a nonbinding resolution (EU Resolution 2014/2841), urging EU members to exclude Russia from the SWIFT system; SWIFT objected to the resolution, reiterating its commitment to neutrality.

In March 2022, the European Union, in consultation with Canada, Japan, the United Kingdom, and the United States agreed to remove some Russian and Belarusian banks from SWIFT (Council Regulations 2022/345 and 398). SWIFT complied with the new EU regulation and, on March 12, 2022, it disconnected seven Russian and three Belarusian banks and their subsidiaries from its network. Three more Russian banks, one more Belarusian bank, and their subsidiaries were disconnected in June 2022.

The Emergence of SWIFT Competitors

The use of the SWIFT system as a tool for financial sanctions by the European Union, the United Kingdom, Canada, and the United States has encouraged other large countries around the world to consider building systems of their own. None of these alternatives has yet been especially successful, but their short-run goal may just be to set up a backup system both to gain expertise in the underlying technology involved and in case their access to SWIFT is threatened in the future.

In 2014, following the political pressures to disconnect Russia from SWIFT, Russia developed its own financial messaging system, SPFS (System for Transfer of Financial Messages). SPFS can transmit messages in the SWIFT format, and more broadly messages based on the ISO 20022 standard, as well as free-format messages. More than 400 banks have already connected to SPFS, most of them Russian or from former Soviet Republics. A few banks from Germany, Switzerland, France, Japan, Sweden, Turkey, and Cuba are also connected. By April 2022, the number of countries with financial institutions using SPFS had grown from 12 to 52, at which point the Central Bank of Russia decided not to publish the names of SPFS users. Due to its limited scale, SPFS mainly processes financial messages within Russia; in 2021, roughly 20 percent of all Russian domestic transfers were done through SPFS, with the Russian central bank aiming to increase this share to 30 percent by 2023 (Shagina 2021).

In 2019, following Iran's loss of access to SWIFT caused by the US threat of sanctions, France, Germany, and the United Kingdom developed the Instrument in Support of Trade Exchanges (INSTEX), a special-purpose vehicle with the mission of facilitating non-SWIFT transactions with Iran. INSTEX was joined by other EU nations and made available to all member states. Although its use is limited

to humanitarian purposes, it provides an example of parallel system to SWIFT set up by countries to side-step the threat of sanctions by another country. Note that INSTEX is not just a messaging system, but rather a clearinghouse that allows payments between Europe and Iran; payments are netted within the system, and direct payments between Iran and the European Union happen only if there are import-export imbalances. Although the system is operational, it has been largely unused since its setup. Indeed, the first INSTEX transaction did not happen until March 2020, covering the import of medical equipment to combat the COVID-19 outbreak in Iran.

Sometimes a country decides to set up a parallel system to SWIFT for purposes that reach beyond immediate concerns over sanctions. In 2015, the People's Bank of China launched the Chinese Cross-Border Interbank Payment System (CIPS) with the purpose of supporting the use of the renminbi in international trade and international financial markets. In contrast to SWIFT, but similar to INSTEX, CIPS is not only a messaging system but also offers payment clearing and settlement services for cross-border payments in renminbi. It started with 19 direct participants and 176 indirect participants from 50 countries; at the end January 2022, there were 1,280 participants from 103 countries. Among the direct participants, eleven are foreign banks, including large banks from the United States and other developed countries. The system is overseen and backed by People's Bank of China. Similarly to Russia's SPFS, CIPS uses the SWIFT industry standard for syntax in financial messages. Indirect participants can obtain services provided by CIPS through direct participants. In 2021, CIPS processed millions of transactions for a total value of around 80 trillion yuan ($12.7 trillion).

Several Russian banks are connected to China's CIPS as indirect participants, which facilitate Russia's business in renminbi, whereas only one Chinese bank is connected to Russia's SPFS. The presence of SPFS and CIPS allows participant institutions to interact with Russian banks, even if these banks are disconnected from SWIFT. However, although CIPS has more participants than SPFS, its overall presence is not comparable to that of SWIFT; the CIPS payment volume is about 0.3 percent of the size of SWIFT. Most of China's CIPS transactions still actually use the SWIFT network, as many firms do not have access to a separate CIPS terminal (Yeung and Goh 2022). Therefore, it may be hard for CIPS to become a viable substitute to the SWIFT network in the near future.

Finally, since 2001, India has developed its own secure messaging network for financial transactions, the Structured Financial Messaging System (SFMS), which allows inter- and intra-bank messaging within India. Similarly to China's CIPS and Russia's SPFS, SFMS supports the ISO 20022 standard and is therefore compatible with SWIFT. A fundamental difference is that SFMS is a purely domestic messaging system. As a result, India does not bypass SWIFT for international transactions. In October 2019, Russian, Chinese, and Indian news media reported that these countries plan to link their respective systems together (Wong and Nelson 2021), but the extent to which this has actually happened is not clear.

The Bottom Line

Financial sanctions have been often used in international relationships, especially by Western countries, with their importance increasing over the recent decades. Sanctions that restrict access to the institutions and infrastructure supporting international payments, such as the SWIFT network, are particularly disruptive. Any kind of cross-border economic activity, be it financial or real, requires access to the international payment system. Recently, some countries have invested in creating alternative systems to allow cross-border payments without relying on institutions based in the West, such as SWIFT. While these alterative systems are currently limited in scope, over time they could meaningfully reduce the effectiveness of restricting access to the existing infrastructure of cross-border payments based in the West.

■ *We thank Catherine Huang, Stone Kalisa, and Sergio Olivas for outstanding research assistance. We thank Richard Charlton, Heidy Medina, Brett Phillips, and Janine Tramontana for helpful comments. The views in this paper should not be interpreted as reflecting the views of the Federal Reserve Bank of New York or the Federal Reserve System.*

References

Amicelle, Anthony. 2011. "The Great (Data) Bank Robbery: Terrorist Finance Tracking Program and the 'SWIFT Affair.'" Unpublished.

Bank for International Settlements. 2016. "Correspondent Banking." Unpublished.

Bank for International Settlements. 2020. "CPMI Correspondent Banking Chartpack." Unpublished.

Bech, Morten Linnemann, and Jenny Hancock. 2020. "Innovations in payments." *BIS Quarterly Review* 23 (1): 21–36.

Becker, Charles M. 1988. "The Impact of Sanctions on South Africa and Its Periphery." *African Studies Review* 31 (2): 61–88.

Chang, Semoon. 2006. "The Saga of U.S. Economic Sanctions against North Korea." *The Journal of East Asian Affairs* 20 (2): 109–40.

Cipriani, Marco, Linda S. Goldberg, and Gabriele La Spada. 2023. "Replication data for: Financial Sanctions, SWIFT, and the Architecture of the International Payment System." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E183945V1.

Connorton, Patrick M. 2007. "Tracking Terrorist Financing Through SWIFT: When U.S. Subpoenas and Foreign Privacy Law Collide." *Fordham Law Review* 76 (1): 283–322.

Crawford, Neta C., and Audie Klotz. 1999. *How Sanctions Work Lessons from South Africa*. Houndmills: Macmillan Press.

De Goede, Marieke. 2012. "The SWIFT Affair and the Global Politics of European Security." *Journal of Common Market Studies* 50 (2): 214–30.

European Commission. 2006. *The EU's Relations with Burma/Myanmar*. European Commission. Brussels: European Commission.

Financial Action Task Force. 2016. "Guidance on Correspondent Banking Services." Paris: FATF. https://www.fatf-gafi.org/publications/fatfrecommendations/documents/correspondent-banking-services.html.

Felbermayr, Gabriel, Aleksandra Kirilakha, Constantinos Syropoulos, Erdal Yalcin, and Yoto Yotov. 2020. "The Global Sanctions Data Base." *European Economic Review* 129 (C).

Francioni, Francesco, and Federico Lenzerini. 2003. "The Destruction of the Buddhas of Bamiyan and International Law." *European Journal of International Law* 14 (4): 619–51.

Galati, Gabriele. 2002. "Settlement Risk in Foreign Exchange Markets and CLS Bank." *BIS Quarterly Review* 10 (4): 55–65.

Ghufran, Nasreen. 2001. "The Taliban and the Civil War Entanglement in Afghanistan." *Asian Survey* 41 (3): 462–87.

Giumelli, Francesco, and Paul Ivan. 2013. "The Effectiveness of EU Sanctions an Analysis of Iran, Belarus, Syria, and Myanmar (Burma)." Unpublished.

Grolleman, Dirk Jan, and David Jutrsa. 2017. "Understanding Correspondent Banking Trends: A Monitoring Framework." IMF Working Papers 2017/216.

Hardister, Angela D. 2002. "Can We Buy a Peace on Earth: The Price of Freezing Terrorist Assets in a Post-September 11 World." *North Carolina Journal of International Law* 28 (3): 605–61.

Hefti, Chantal, and E. Staehelin-Witt. 2011. "Economic Sanctions against South Africa and the Importance of Switzerland." Unpublished.

Helwege, Ann. 1989. "Three Socialist Experiences in Latin America: Surviving US Economic Pressure." *Bulletin of Latin American Research* 8 (2): 211–34.

Hufbauer, Gary Clyde, Jeffrey Schott, and Barbara Oegg. 2001. *Using Sanctions to Fight Terrorism*. Washington, DC: Peterson Institute for International Economics.

Kirilakha, Aleksandra, Gabriel J. Felbermayr, Constantinos Syropoulos, Erdal Yalcin, and Yoto V. Yotov. 2021. "The Global Sanctions Data Base: An Update that Includes the Years of the Trump Presidency." In *The Research Handbook of Economic Sanctions*. edited by Peter A. G. van Bergeijk, 62–106. Cheltenham: Edward Elgar Publishing.

Koppel, Johannes. 2011. *The SWIFT Affair*. Geneva: Graduate Institute Publications.

Levy, Philip I. 1999. "Sanctions on South Africa: What Did They Do?" *American Economic Association* 89 (2): 415–20.

Livingstone, Grace. 2009. *America's Backyard: The United States Latin America from the Monroe Doctrine to the War on Terror*. London: Zed Books.

Martin, Michael F. 2012. "U.S. Sanctions on Burma." Unpublished.

Miller, Rena S. 2022. "Overview of Correspondent Banking and 'De-Risking' Issues." Unpublished

Olson, Richard Stuart. 1979. "Economic Coercion in World Politics: With a Focus on North-South Relations." *World Politics* 31 (4): 471–94.

Petras, James, and Morris Morley. 1975. *The United States and Chile: Imperialism and the Overthrow of the Allende Government*. New York: Monthly Review Press.

Petras, James, and Morris Morley. 1978. "On the U.S. and the Overthrow of Allende: A Reply to Professor Sigmund's Criticism." *Latin American Research Review* 13 (1): 205–21.

Rice, Tara, Goetz von Peter, and Codruta Boar. 2020. "On the Global Retreat of Correspondent Banks." *BIS Quarterly Review* 23 (1): 37–52.

Scott, Susan V., and Markos Zachariadis. 2012. "Origins and Development of SWIFT, 1973–2009" *Business History* 54 (3): 462–82. https://www.bis.org/publ/qtrpdf/r_qt2003g.htm.

Scott, Susan V., and Markos Zachariadis. 2014. *The Society for Worldwide Interbank Financial Telecommunication (SWIFT): Cooperative Governance for Network Innovation, Standards, and Community*. London: Routledge.

Sechrist, Michael. 2010. "Cyberspace in Deep Water: Protecting Underseas Communication Cables by Creating an International Public-Private Partnership." Unpublished

Shagina, Maria. 2021. "How Disastrous Would Disconnection from SWIFT Be for Russia?" Carnegie Endowment for International Peace." Carnegie Endowment for International Peace, May 28th. https://carnegiemoscow.org/commentary/84634.

Sigmund, Paul E. 1974. "The 'Invisible Blockade' and the Overthrow of Allende." *Council on Foreign Relations* 52 (2): 322–40.

SWIFT. 2020. "SWIFT Annual Review: 2020." Unpublished.

SWIFT. 2021. RMB Tracker January 2021 Monthly Reporting and Statistics on Renminbi (RMB)

Progress towards Becoming an International Currency." https://www.swift.com/our-solutions/compliance-and-shared-services/business-intelligence/renminbi/rmb-tracker/rmb-tracker-document-centre?page=2.

Wong, Liana, and Rebecca M. Nelson. 2021. "International Financial Messaging Systems." Unpublished.

World Bank. 2022. "An Analysis of Trends in Cost of Remittance Services: Remittance Prices Worldwide Quarterly." Washington, DC: The World Bank Group.

Yeung, Raymond, and Khoon Goh. 2022. "Petroyuan Will Not Bring About a Regime Shift Soon." Unpublished.

Journal of Economic Perspectives—Volume 37, Number 1—Winter 2023—Pages 53–76

Monetary Policy When the Central Bank Shapes Financial-Market Sentiment

Anil K Kashyap and Jeremy C. Stein

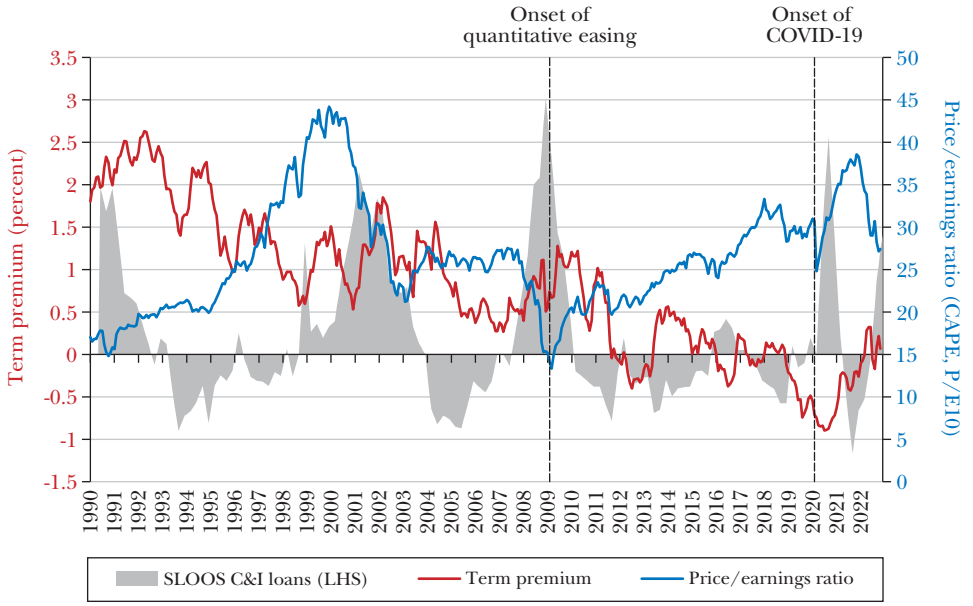
In the years since the global financial crisis of 2008–2009, the conduct of monetary policy has changed markedly. As central banks sought to stimulate the macroeconomy with lower interest rates, but then bumped up against the zero lower bound on short-term policy rates, they began to experiment with other tools, most notably by buying large amounts of financial assets—that is, by engaging in quantitative easing or "QE"—to raise the prices of these assets and lower their yields.

While Ben Bernanke was Chair of the Federal Reserve, he joked that "the problem with QE is that it works in practice but not in theory" (Bernanke 2014). Figure 1 displays some broad evidence consistent with the works-in-practice view. The graph shows three prominent measures of financial conditions. The first is the ratio of stock prices relative to an average of corporate earnings over the prior ten years. This ratio is driven in part by the rate at which market participants discount earnings, with higher multiples indicating lower discount rates, all else being equal. The second measure is based on survey responses from US banks and indicates the net percentage of banks that are tightening their lending standards for new commercial and industrial loans. This variable is a proxy for whether bank credit is becoming harder or easier to obtain, with lower values indicating easier credit conditions. Finally, the third measure is an estimate of the "term premium" on a

■ *Anil K Kashyap is Stevens Distinguished Service Professor of Economics and Finance, University of Chicago Booth School of Business, Chicago, Illinois. Jeremy C. Stein is Moise Y. Safra Professor of Economics, Harvard University, Cambridge, Massachusetts. Both authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are anil.kashyap@ChicagoBooth.edu and jeremy_stein@harvard.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.1.53.

Figure 1
Evolution of Stock Prices, Treasury Term Premiums, and Bank Lending Standards During the Quantitative Easing Era



Source: For data sources and the underlying calculations, see Kashyap and Stein (2023).
Notes: The blue solid line plots Robert Shiller's cyclically adjusted price-earnings ratio (CAPE). The red solid line plots estimated values of the Treasury term premium, based on the methodology of Kim and Wright (2005). The grey shaded area shows the net percentage of domestic banks tightening standards for commercial and industrial (C&I) loans to large and middle-market firms, from the Federal Reserve Senior Loan Officer Opinion Survey (SLOOS). For graphical convenience, this SLOOS series is rescaled to have the same approximate range as the term premium series. The sample period runs from 1990 to 2022.

ten-year zero-coupon Treasury bond. One component of a ten-year bond's yield reflects the so-called expectations hypothesis—the idea being that the ten-year rate should resemble the expected average of short-term rates over the next ten years. The other component of the longer-term yield, the term premium, is the additional compensation above and beyond this expectational piece. When the term premium is high, it means that investors are demanding more compensation for the risk associated with investing for longer periods (and vice-versa).

Figure 1 shows that, according to all three of these measures, financial conditions loosened considerably with the initiation of quantitative easing by the Federal Reserve in 2009 and remained relatively loose for the next decade, up until the onset of the COVID-19 pandemic in March 2020. In particular, price-earnings ratios on stocks were generally rising during this period, bank lending standards were relaxed, and term premiums on US Treasury bonds were well below their historical averages. The low values of Treasury term premiums may not be all that surprising

given that the Fed was directly buying Treasury securities, but the movement in the other measures is not mechanically linked to the Fed's purchase programs.

As Bernanke's comment suggests, during the early days of quantitative easing, practice was running ahead of theory, in the sense that conventional macroeconomic models did not offer a clear explanation for *why* central bank asset purchases should have such widespread effects on asset prices. In recognition of this gap in understanding, both theoretical and empirical research began to focus increasingly on a variety of institutional and behavioral frictions absent from traditional models in an effort to better understand the mechanisms of quantitative easing. As we discuss in detail below, once researchers began to take these frictions more seriously, it became clear that they not only help to explain the workings of quantitative easing, they also offer a new and powerful way of thinking about the channels of influence of plain-vanilla conventional monetary policy—that is, policy implemented solely through changes in short-term interest rates such as the federal funds rate.

In particular, it now appears clear that both conventional and unconventional monetary policy actions gain much of their traction over the real economy by influencing a range of risk premiums in financial markets, where the risk premium on an asset is the expected return that an investor can expect to earn above and beyond the safe rate on a government bond of comparable maturity. When risk premiums are low, this can be thought of as a time when investors are either relatively risk tolerant, or relatively optimistic, so they drive asset prices up, and hence push future expected returns down. The work we review below, which is also discussed in the companion paper in this symposium by Bauer, Bernanke, and Milstein, documents that an easing of monetary policy tends to reduce risk premiums on a range of financial assets, including stocks, Treasuries, corporate bonds, and foreign exchange. Similarly, easy monetary policy tends to make banks more willing to accept a lower return for taking the credit risk associated with the loans they make.

These policy-induced movements in risk premiums, like movements in risk premiums more generally, tend to be temporary, meaning that an increase in asset prices spurred by central bank action is typically reversed in the months or years that follow. One way to summarize these findings is to say that central banks have a broad ability—through both their conventional and unconventional policies—to influence financial-market "sentiment," which we use as a synonym for the time-varying risk premiums on both traded securities and intermediated loans. For example, when credit spreads are compressed because bond prices have been bid up and the objective expected return to bearing credit risk is unusually low, we will say that credit-market sentiment is elevated. In this usage, elevated sentiment can reflect either a change in a rational investor's attitude toward risk—that is, a willingness to knowingly accept lower returns—or behavioral mistakes of various sorts that lead investors to be overly optimistic about future outcomes. For much of what follows, we can be agnostic as to which of these two mechanisms is at work. Either way, this channel of monetary-policy transmission is very different than what is envisioned in traditional textbook models.

The central thesis of this paper is that once one appreciates that monetary policy achieves much of its effectiveness through its impact on financial-market

sentiment, one may think quite differently about certain issues in the conduct of policy. To see why, it helps to connect to a second strand of recent work, which documents the importance of what we call a "credit-bites-back" effect in homage to the seminal paper of Jordà, Schularick, and Taylor (2013). In brief, this literature finds that following periods of rapid credit growth, and especially when asset prices are elevated and risk premiums are compressed (that is, when sentiment is running high), the likelihood of a recession or a financial crisis significantly increases.

Taken together, these two lines of research suggest a potentially important tradeoff facing monetary policymakers. Accommodative policy can be quite powerful in raising asset prices and spurring aggregate demand, even if short-term interest rates are stuck near zero; this is the upside of the central bank's ability to stoke market sentiment with tools that affect risk premiums. However, this power comes with a potential downside as well, because elevated sentiment today is likely to reverse eventually, and in doing so, it may increase the odds of a recession at some later date. As we argue below, this tension becomes all the more pronounced when financial regulation is by itself unable to fully contain the credit-bites-back risks put into play by monetary policy.

New Theories about Monetary Policy Transmission

In canonical New Keynesian accounts of monetary policy transmission, time-variation in financial-market risk premiums does not play a meaningful role (for traditional models, see Woodford 2003; Galí 2008, 2018). In these models, when the central bank cuts the short-term nominal rate, the assumption of price stickiness implies that it also lowers the short-term real interest rate. If monetary policy changes are persistent, there will be an associated impact on longer-term real rates as well; these in turn will influence consumption and investment decisions. This story can largely be told in a world where all risk premiums are constant over time.

A similar observation applies to other familiar accounts of the monetary transmission mechanism, such as the "bank lending channel" (Kashyap and Stein 2000; Drechsler, Savov, and Schnabl 2017). Here, an easing of monetary policy allows banks to raise additional deposits and expand their lending. This could be, for example, because in a low-interest-rate environment, banks do not have to compete as aggressively for retail deposits with higher-yielding alternatives such as money-market funds. In this theory, what changes for banks as monetary policy varies is not their risk tolerance, but rather their liquidity position, and hence their ability to finance their lending activity. The broader macro literature on the financial accelerator, as summarized by Bernanke, Gertler, and Gilchrist (1999), relies on the idea that as collateral values increase in good times, households and firms are able to borrow more and hence expand economic activity, but it also does not emphasize time-varying risk premiums as a central factor in policy effectiveness.

By contrast, a body of recent work has put changes in investor and intermediary willingness to bear risk front and center in its account of monetary policy. We

begin here with a brief discussion of the underlying theories; in the next section, we review four new facts that support these theories.

Changes in central-bank policy rates might affect the willingness of investors and intermediaries to take risk, and hence the risk premiums on a range of financial assets, through several channels. For example, one idea is based on the premise that investors face a sustainable spending constraint and can only consume the expected returns from their wealth—that is, investors do not wish to run down their wealth over time. This assumption seems to capture the behavior of endowments and sovereign wealth funds, as well as perhaps that of some individual retirees. Campbell and Sigalov (2022) build this assumption into a neoclassical model of consumption and portfolio choice featuring an infinitely-lived investor. The presence of a sustainable spending constraint naturally generates "reaching for yield" behavior: as the real interest rate falls, the investor tends to increase their portfolio's allocation to risky assets in an effort to partially maintain their level of current consumption.

Another approach emphasizes the ways in which agency or regulatory frictions can distort intermediary behavior (in the spirit of Rajan 2005; Borio and Zhu 2012). For example, Hanson and Stein (2015) build a model in which a set of intermediaries such as commercial banks care about maintaining their accounting income in the face of interest-rate cuts. This leads the intermediaries to take on more "duration risk" at such times—that is, to be more willing to hold longer-term bonds—which in their model puts downwards pressure on the term premium between long- and short-term debt. Chodorow-Reich (2014) and Di Maggio and Kacperczyk (2017) argue that periods of low interest rates may lead money-market funds to take more risk in order to cover their fixed costs and sustain their profit margins.

Drechsler, Savov, and Schnabl (2018) and Acharya and Naqvi (2019) take a somewhat different route, noting that accommodative monetary policy gives banks easier access to cheap liquidity, which serves an insurance role: they can afford to take on more risk without worrying as much about whether this additional risk might cause a disruptive liquidity shortfall.

A separate group of models sets aside these kinds of constraints and frictions, and instead focuses either on how monetary policy can affect the distribution of wealth or on explicitly behavioral factors. For example, Kekre and Lenel (2022) highlight the importance of heterogeneity in households' risk tolerance and argue that an interest-rate cut redistributes wealth towards more risk-tolerant households, thereby increasing aggregate risk appetite. Adopting a more behavioral perspective, Lian, Ma, and Wang (2019) find in randomized experiments that people exhibit a stronger preference for risky assets when the risk-free rate is lower, which they interpret as evidence that psychological mechanisms, such as reference points and salience, affect investor risk-taking in an important way. In another behavioral model, due to Fontanier (2022), a rate cut that initially raises asset values for purely fundamental discounted-cashflow reasons also causes investors who extrapolate from past price increases to become overly enthusiastic about future prospects, thereby causing an eventual overshoot of valuations.

Supporting Evidence on Monetary Policy and Risk Premiums

We review four patterns of facts that confirm the predictions of the above theories related to risk-taking. In particular, we discuss evidence on how changes in the stance of monetary policy influence: (1) the term premiums on government bonds; (2) stock market risk premiums; (3) the pricing of credit risk in both corporate bonds and in bank lending terms; and (4) foreign exchange risk premiums. In each case, looser monetary policy, whether it is initiated by interest rate changes or unconventional means such as quantitative easing, leads to lower risk premiums and hence easier financial conditions.

Fact 1: Treasury Term Premiums

Standard discussions of quantitative easing, like Bernanke (2020), point to its impact on the risk premiums of those specific assets that are being purchased by the central bank. It might not be terribly surprising, for example, if large-scale Fed purchases of long-term Treasury bonds lowered their yields relative to short-term interest rates, and hence compressed Treasury term premiums. What may be somewhat more surprising is the finding that even when monetary policy is implemented conventionally, with changes only in the short-term policy rate and the Fed not adding to its holdings of Treasury bonds, there is nevertheless a strong impact of monetary policy on Treasury term premiums.

One illustration of this pattern comes from Hanson and Stein (2015), who study the high-frequency reaction of real interest rates—as captured by the interest rates on Treasury inflation-protected securities—to monetary policy announcements. They find that monetary innovations have a surprisingly large effect on real rates far in the future. For example, if the two-year nominal Treasury yield goes up by 25 basis points in the immediate wake of a monetary policy announcement by the Federal Open Market Committee, this is associated with an 11 basis-point increase in the ten-year forward real rate. Hanson and Stein argue that this increase in the distant-forward real rate is unlikely to reflect a change in the expected path of short-term real rates at such a long horizon—which would require prices to be counterfactually sticky for an extremely long time—but rather a change in the Treasury term premium. In support of this point, they demonstrate that those movements in forward rates that occur on dates when the Federal Open Market Committee makes a policy announcement tend to largely mean revert over the next twelve months. This reversal effect is also suggestive of a change in risk premiums.

In a similar vein, Hanson, Lucca, and Wright (2021) find that since 2000, increases in short-term Treasury rates are associated with strong, yet temporary, upwards pressure on term premiums. They build a model in which changes in short-term interest rates trigger "rate-amplifying" shifts in the demand for long-term bonds, which might come from investors who either extrapolate recent changes in short-term interest rates, or who reach for yield when short rates fall.

Fact 2: Stock Market Risk Premiums

If the stock market reacts to monetary policy surprises, this can create another channel for monetary policy transmission. Bernanke and Kuttner (2005), using an event-study approach, find that a surprise cut of 25 basis points in the federal funds rate target is associated with a contemporaneous increase in the value of the stock market of about 1 percent. Perhaps more interestingly for our purposes, they show that the vast majority of the stock-price increase—on the order of 80 percent—is due to a change in the expected excess return, or risk premium, in the stock market. Concretely, they document that the initial upward spike in stock returns is followed by a period of abnormally low returns; that is, the boost to stock prices associated with a surprise monetary easing is in large part transitory and is eventually mostly reversed. In this symposium, Bauer, Bernanke, and Milstein show that these results continue to hold when the sample is updated through 2022 and when several different measures of monetary policy shocks are considered (see also Cieslak and Pang 2021). They also add another complementary result, namely that looser monetary policy additionally reduces the volatility of stock prices. These patterns are exactly what one would expect to find if the monetary-policy innovation led to an increase in investor risk tolerance.

Fact 3: Credit Spreads and Bank Lending Terms

The risk premium on corporate credit—that is, the expected return differential between risky corporate bonds and safe Treasury bonds—is one of the most important risk premiums that monetary policy can affect, given that risk premiums on corporate credit have been documented to have powerful effects on real economic activity (Gilchrist and Zakrajšek 2012; López-Salido, Stein, and Zakrajšek 2017). However, inference in this case is somewhat trickier than for the Treasury market and the stock market. Corporate bonds are less liquid, and less actively traded than stocks or government bonds, and so may reprice less promptly in the immediate aftermath of a meeting of the Federal Open Market Committee than do other securities. If so, a high-frequency event study looking at the hours just before and after a monetary policy announcement would be biased away from finding an effect of monetary policy on corporate credit spreads, especially if these spreads are measured directly based on the difference in corporate yields and faster-adjusting Treasury yields.

One response to this challenge is to look at longer-horizon effects. In this spirit, Gertler and Karadi (2015) use a vector autoregression to estimate the dynamic impact, at monthly frequency, of monetary-policy surprises on the Gilchrist and Zakrajšek (2012) "excess bond premium." The excess bond premium can be thought of as that portion of the credit spread that is not accounted for by expected default losses, and it therefore maps very closely into the concept of a credit-risk premium. Gertler and Karadi (2015) find that a monetary surprise that reduces the one-year Treasury bill rate by 25 basis points compresses the excess bond premium by 10 basis points in the first month. This effect persists for about eight months, and then is gradually reverted away, again consistent with the behavior of a transitory risk premium.

By its nature, however, this longer-horizon approach is inevitably more sensitive to the precise details of the econometric specification and the identifying assumptions used in estimation; as such, it lacks the appealing transparency and robustness of a high-frequency event study. Bauer and Swanson (2022) provide a detailed treatment of these issues. Interestingly, with their preferred approach to identification they find an even stronger effect of monetary policy surprises on the excess bond premium than do Gertler and Karadi (2015), although they are careful to highlight the sensitivity of these results to alternative specifications.

Another approach to address how monetary policy affects the pricing of credit risk is to revert back to the high-frequency event-study methodology, but to look at the spreads on credit default swaps instead of corporate bonds. Credit default swaps are a financial contract that allows the buyer of a bond to purchase insurance against the risk of the bond defaulting. The market for credit default swaps tends to be more liquid than the underlying bonds, and to have prices that adjust more rapidly, so they may be better-suited to a high-frequency approach. Indeed, using a methodology similar to Hanson and Stein (2015), Palazzo and Yamarthy (2022) find that, in the short window around a monetary policy announcement, a 25 basis-point increase in the two-year Treasury yield is associated with a 7 basis-point average increase in firm-level spreads in credit default swaps. They also uncover noteworthy heterogeneity in the response, with a larger effect being seen in the set of riskier firms that had higher spreads in their credit default swaps before the policy announcement.

Of course, when one thinks about the pricing of credit risk, it is important to go beyond the corporate bond market and also to consider bank lending. One might naturally expect some integration between the pricing of credit risk across corporate bonds and bank loans; for example, such a conjecture is consistent with the relatively high correlation between corporate credit spreads and bank lending terms as reported in the Federal Reserve's Senior Loan Officer Opinion Survey.[1]

As it turns out, an easing of monetary policy does in fact seem to lead banks to loosen their credit standards and take on more credit risk. For example, Paligorova and Santos (2017) use data on syndicated corporate loans from Dealscan to show that when short-term interest rates are low, there is a reduced sensitivity of the spread that a firm is charged on its loans to a measure of its fundamental credit risk; in other words, there is a lower cross-sectional price of credit risk. In a similar vein, Dell'Ariccia, Laeven, and Suarez (2017) exploit supervisory data from the Federal Reserve to look at how banks' internal risk ratings on newly originated loans vary with the stance of monetary policy. They find that when the policy rate declines, banks extend more credit to riskier borrowers. This is true even when they restrict the set of loans only to those that are new and not made under commitment, so that this choice is clearly discretionary. Maddaloni and Peydró (2011) use loan officer

[1] For example, over the period 1996:4–2022:2, the correlation in levels between the high-yield credit spread and a measure of easing of credit terms from the Federal Reserve's Senior Loan Officer Opinion Survey is –0.51. There is also a strong correlation between the opinion survey and corporate bond issuer quality, as noted by Greenwood and Hanson (2013).

survey data from both the United States and the euro area to document that times of low policy rates are associated with generally laxer lending standards.

These sorts of results hold across a range of other countries. Using credit registry data from Spain, Jimenez et al. (2014) find that when interest rates drop, the amount of lending to firms with bad credit histories (or future impending losses) rises relative to loans made to more creditworthy firms. They also show that this effect is more pronounced for loans made by weakly-capitalized banks than for those made by well-capitalized ones. Using data from Bolivia—a largely dollarized economy where monetary policy changes are exogenously transmitted from the United States—Ioannidou, Ongena, and Peydró (2015) show that a lower federal funds rate leads to relatively more lending by Bolivian banks to borrowers with worse credit histories, lower internal credit ratings, and who display poorer post-loan performance.

Thus, whether through banks or via bond markets, an important part of what happens when the central bank eases monetary policy is that the risk premium on corporate credit declines. Holding fixed both borrowers' creditworthiness and loan demand, we would expect to see lower policy interest rates followed by an expansion in overall credit creation, and one that is tilted towards higher-risk firms.

Fact 4: Foreign Exchange

Both long-term bonds and exchange rates are exposed to a common primary risk factor—namely, changes in the stance of monetary policy. With this observation in mind, Greenwood et al. (forthcoming) and Gourinchas, Ray, and Vayanos (2022) argue that there is likely to be a close correlation between bond market term premium differentials across countries on the one hand, and exchange-rate risk premiums on the other. Greenwood et al. (forthcoming) provide supporting evidence, showing for example that if the Federal Reserve undertakes a round of quantitative easing, it both reduces the term premium on US Treasury securities relative to term premiums in other countries, and also weakens the value of the US dollar—but only for a time, so that the dollar subsequently tends to appreciate by an abnormal amount going forward against other currencies. This finding offers yet another example of central-bank policy gaining additional traction to stimulate output in the short term by virtue of its ability to influence risk premiums.

Evidence on the Credit-Bites-Back Mechanism

With these four facts about monetary policy and risk premiums in hand, we now turn to the body of work that studies the credit-bites-back mechanism. Broadly speaking, this work highlights two other patterns of facts. First, if one looks at quantity data that captures the growth of aggregate credit, then at relatively low frequencies, rapid growth in credit tends to portend adverse macroeconomic outcomes, be it a financial crisis or some kind of more modest slowdown in activity. Second, elevated levels of financial-market sentiment—especially indicators which signal that the expected returns to bearing credit risk are low—also tend to carry

negative information about future economic growth, above and beyond that the information present in credit-quantity variables. Thus, the overall picture is that credit booms, especially those associated with both rapid increases in the quantity of credit and also aggressive pricing of credit risk, tend to end badly. The summary that follows draws heavily on Stein (2021).

With respect to the quantity-oriented evidence, some of the most influential research comes from Schularick and Taylor (2012) and Jordà, Schularick, and Taylor (2013). In the former, they study 14 developed countries over the period 1870–2008 and find that the growth of bank loans in the preceding five years is associated with a significantly increased probability of a financial crisis. In a similar spirit, Mian, Sufi, and Verner (2017) also focus on a quantitative measure of credit expansion, in this case the ratio of household credit to GDP. Using a sample of 30 mostly advanced economies and a panel running from 1960 to 2012, they find large negative effects of credit booms on future output: a one-standard-deviation increase in household debt to GDP over a three-year interval leads to a 2.1 percent decline in GDP over the following three years. Notably, these results reflect not just occurences of extreme financial crises, but are also driven by more moderate noncrisis recessions and slowdowns. Sufi and Taylor (2021) provide an excellent summary of the recent research on financial crises.

Turning to the connection between credit-market sentiment and future growth, López-Salido, Stein, and Zakrajšek (2017) investigate the role of sentiment in a US sample running from 1929 to 2015. To do so, they build on the work of Greenwood and Hanson (2013), who show that when credit spreads are narrow, and when the share of high-yield (or "junk bond") issuance in total corporate bond issuance is high, the expected returns to bearing credit risk are predictably low, and sometimes even negative—in other words, narrow credit spreads and an above-average high-yield share, taken together, are indicative of elevated credit-market sentiment. López-Salido, Stein, and Zakrajšek (2017) then show that exuberant credit-market sentiment in a given year t is associated with a decline in economic activity in years $t + 2$ and $t + 3$. Underlying this result is the existence of predictable mean reversion in credit-market conditions. When credit risk is aggressively priced, spreads subsequently widen. The timing of this widening is closely tied to the onset of a contraction in economic activity, one in which the pain is felt disproportionately by firms with lower credit ratings. Exploring the mechanism, they find that buoyant credit-market sentiment in year t also forecasts a change in the composition of external finance: net debt issuance falls in year $t + 2$ while net equity issuance increases, consistent with the reversal in credit-market conditions leading to an inward shift in credit supply.

This focus on the impact of investor sentiment on future economic outcomes is extended by Kirti (2018) in a sample encompassing 38 countries. His key finding concerns the *interaction* of growth in the quantity of credit with credit-market sentiment, where he follows Greenwood and Hanson (2013) and proxies for sentiment with the high-yield share of bond issuance. In particular, following strong credit growth, economic growth in the following three years is roughly 1.1 percent slower per year. However, if this increase in the quantity of credit is accompanied by a

two-standard-deviation increase in the high-yield share, growth over the next three years slips by a further 0.8 percent per year. Krishnamurthy and Muir (2020) present related findings, using a panel that goes back 150 years and covers 19 countries.

Greenwood et al. (2022) also analyze the interaction between credit growth and asset prices, using a panel of 42 countries over the period 1950 to 2016. They examine what happens when a country enters a vulnerable "Red Zone," characterized by business credit growth over the prior three years in the top quintile of the distribution, and stock returns over the same window in the top tercile. For countries in the Red Zone, the probability of a financial crisis rises dramatically—from a normal-times value of 7 percent over a three-year horizon to over 40 percent.

A related set of papers uses quantile regressions to explore how changing financial conditions affect not just mean or median outcomes, but the *full distribution* of real activity over a subsequent time period. For example, Adrian, Boyarchenko, and Giannone (2019) focus on the US evidence, while Adrian et al. (2022) also look at data from Australia, Canada, Switzerland, Germany, Spain, France, Great Britain, Italy, Japan, and Sweden. The general picture that emerges in both studies is that it is the lower tail of GDP growth—for example, the fifth percentile—that seems especially vulnerable in the two to three years following an easing of financial conditions. In other words, loose financial conditions seem to raise the downside risks to real activity, while having a weaker effect on the upper tail of the distribution.

What specific measures of financial conditions are most relevant in this sort of predictive exercise? Using US data, Carpenter et al. (2022) find that proxies for credit supply such as loan spreads or debt levels are more informative for downside risks to the economy than variables relating to equity markets or exchange rates. The idea that tracking the pricing of credit risk is especially important in this context echoes the findings of López-Salido, Stein, and Zakrajšek (2017), among others.

We believe that the above-discussed evidence is quite compelling in establishing two propositions: (1) accommodative monetary policy leads to reductions in risk premiums generally, and in credit risk premiums in particular; and (2) rapid credit growth and compressed credit risk premiums increase the odds of adverse economic outcomes at a horizon of between two to five years.

However for the purposes of using these empirical findings to draw implications for the conduct of monetary policy, two caveats should be noted. First, as pointed out by Boyarchenko, Favara, and Schularick (2022), there is limited evidence that it is specifically *monetary-policy induced changes* in credit growth and risk premiums—as opposed to changes driven by other factors—that create this economic vulnerability. As they note, establishing such a link is challenging, and more research on this specific issue would be valuable. We are going to make the leap and assume that the link is operative in what follows, but the reader should be aware that this presumption is not yet firmly established.

Second, any normative implications for monetary policy hinge on the extent to which the credit-bites-back risks we have identified can be mitigated by financial regulation. A traditional argument is that financial regulation should be the first line of defense against these risks (for example, Bernanke 2015). While agreeing

on the importance of robust financial regulation, Stein (2021) expresses skepticism about its ability to serve as a panacea. He notes that the limitations of financial regulation are likely to vary by jurisdiction, but are particularly acute in countries like the United States, where the majority of corporate credit creation now takes place outside the easier-to-regulate banking sector, and where various political-economy constraints have left policymakers with essentially nothing in the way of time-varying macroprudential tools that can be used to address a sharp deterioration in observed credit standards and quality. With this observation in mind, our implicit assumption in the remainder of the paper is that even after doing the best that one can with existing financial-regulation tools, there still remains—as in the historical data—a meaningful credit-bites-back effect.

A Model of Monetary Transmission via Credit Risk Premiums

In what follows, we describe a bare-bones framework in which one can examine the intertemporal tradeoff that arises when monetary policy influences credit risk premiums and when there is a credit-bites-back effect of the sort documented in the work discussed above. We proceed here by just describing our basic assumptions and conclusions. A more complete analysis of the model appears in the online Appendix. Other models that investigate similar issues are Caballero and Simsek (2020, 2022), Adrian and Duarte (2020), and Fontanier (2022).

The Textbook Case

To keep things simple, we assume that the central bank has no inflation mandate, so that its only responsibility is output stabilization. This assumption can be thought of as capturing a "divine coincidence" world where shocks only come from the demand side of the economy, and so stabilizing output also amounts to stabilizing inflation. A textbook rendition of the so-called IS (investment-saving) curve, which captures the effect of interest rates on spending (also known as "aggregate demand") is given by:

$$y_t = y^* - \gamma(r_t - r^*) + \epsilon_t,$$

where y_t is output at time t, y^* is potential output, r_t is the real interest rate, r^* is the natural rate of interest, and ϵ_t is an aggregate demand shock. In this textbook case it is easy to show that the central bank can stabilize output perfectly period-by-period, by raising (lowering) interest rates the appropriate amount in the face of a positive (negative) demand shock.

Adding Credit Spreads

To capture the financial-market effects we have been discussing, we now add credit spreads to the model and allow monetary policy to influence these spreads. To be clear, although we use the terms "credit" and "credit spreads" for concreteness in what follows, our analysis would apply equally to other risk premiums that

are influenced by monetary policy, such as the stock market risk premium, bank lending spreads, or term premiums in the Treasury market.

With this added bit of realism, the IS curve is modified as follows:

$$y_t = y^* - \gamma((r_t + s_t) - (r^* + s^*)) - \beta(s_t - s_{t-1}) + \epsilon_t,$$

where s_t is the credit spread at time t, and s^* is the steady-state value of the credit spread.

There are two changes to note here: first, what matters for aggregate demand now is not the real interest rate set by the central bank, but a broader notion of financial conditions, given by the current value of $(r_t + s_t)$ relative to its long-run average value of $(r^* + s^*)$. Second, and crucially, there is a "credit-bites-back" term, given by $-\beta(s_t - s_{t-1})$: output is reduced, all else equal, when credit spreads *increase* from the prior period. This might be because an increase in credit spreads impairs the health of financial intermediaries, and financial regulation is inadequate to fully prevent this damage. For example, a bank's capital might be reduced by an erosion of the perceived credit quality of its loan book, and this might in turn compromise its ability to make new loans. Or a corporate bond fund that experiences mark-to-market losses might see substantial outflows of money under management, which would dampen its demand for new bonds.

The time-t credit spread is in turn determined by:

$$s_t = s^* + \theta(r_t - r^*) + \upsilon_t,$$

where the $\theta(r_t - r^*)$ term captures what can be thought of as a reaching-for-yield effect—easy monetary policy tends to depress credit spreads—and υ_t is an exogenous credit-supply shock.

The parameter β is key to creating an intertemporal tradeoff for policy. To see why, suppose $\beta = 0$, so there is no credit-bites-back effect. In this case, output can again be perfectly stabilized in every period with a simple modification of the interest-rate rule. Relative to the simpler textbook case, the interest-rate rule in this case is changed in two ways. First, the policy rate is less responsive to demand shocks. This is because changes in the policy rate have an amplified impact on output, due to the reaching-for-yield effect. Second, policy leans against exogenous movements in financial conditions, as given by υ_t. When credit spreads are relatively low, the policy rate is higher, and vice-versa.

Thus, in this limiting case where $\beta = 0$, and there is no credit-bites-back effect, optimal monetary policy takes account of both exogenous changes in financial conditions, as well as its own impact on these conditions. Note, however, that to do so the central bank must be able to observe the exogenous credit supply shock υ_t precisely, which amounts to being able to separate these temporary shocks to credit conditions from more permanent shifts in steady-state credit spreads, as denoted by s^*. This informational requirement is potentially challenging. Nevertheless, if we provisionally assume that υ_t can be well measured, monetary policy faces no compromises or tradeoffs and is still able to perfectly stabilize output in every period.

This version of the model might be thought of as roughly in line with contemporary central-bank practice, whereby a good deal of attention is paid to financial conditions—and where evidence suggests that the policy rate is indeed set at a lower value, all else equal, when conditions are tight, and vice-versa (Peek, Rosengren, and Tootell 2016; Razzak 2022)—but where the intertemporal tradeoffs associated with policy-induced changes in financial conditions are generally not given explicit consideration, at least not in the formal models used to guide policy.

An Intertemporal Tradeoff

To see how an intertemporal tradeoff can arise, suppose instead that $\beta > 0$, so that a credit-bite-back effect exists. To simplify the exposition, we can focus on a two-period version of the model, where what matters are the policy rates r_1 and r_2 at times 1 and 2, respectively, and the tradeoffs these choices entail. Moreover, we assume that at an earlier time 0, the economy was in steady state, with $r_0 = r^*$, and with $s_0 = s^*$. To simplify even more, we further assume that there are no credit supply shocks at either time 1 or time 2, so that $v_1 = v_2 = 0$. Finally, the most interesting scenario arises when there are persistent recessionary pressures—that is, negative demand shocks—at both dates, and there is a possibility that things may get worse at time 2, to the point that the zero lower bound on interest rates may bind, meaning that the central bank may be unable to restore the economy to full employment at time 2 by cutting interest rates as far as this would require.

A richer model could also allow for other reasons, besides the zero lower bound, why policy might be unable to fully neutralize all relevant shocks to the economy: for example, perhaps the rapid unwinding of a financial bubble has an especially damaging effect on the credit-allocation mechanism. Alternatively, lags in the transmission of policy to the real economy may make it harder to offset negative shocks fully. However, to make our points as simply as possible we set aside these considerations and use the zero lower bound as a catchall for the idea that there may be times when monetary policy cannot perfectly offset all potential damage to the real economy.

In this configuration, we can demonstrate a number of propositions. In particular, if the zero lower bound binds at time 2, then: (1) the optimal policy rate at time 1 is higher than it would be if the zero lower bound were not binding at time 2; (2) output at time 1 is lower than it would be if the zero lower bound were not binding at time 2; and (3) it is no longer optimal for the central bank to offset negative time-1 demand shocks fully.

Intuitively, the central bank fears that if it cuts rates at time 1 enough to offset a negative demand shock fully, it will overheat credit conditions, and this overheating will create a drag on time-2 output that cannot be offset if the zero lower bound binds at time 2. This is the core intertemporal tradeoff that arises in our setting. Moreover, this time-1 timidity in providing accommodation is more pronounced when the anticipated negative demand shock at time 2 is larger in absolute magnitude—or, in a richer setting, when the likelihood of a severe zero lower bound episode is greater.

The upshot is that considerations of financial stability can, in some cases, make the central bank choose to stop short of hitting its full-employment mandate if hitting this mandate would require overheating financial markets to the point that employment in future periods is put at too much risk. Of course, it is well-understood that a central bank might stop short of hitting its full-employment mandate if inflation is running above its target, so tradeoffs of this general sort are familiar to central bankers. What is different in our setting is that the tradeoff is not between full employment today and inflation today, but rather between full employment today and full employment tomorrow, with the potential for financial-market reversals being the link that binds these two items together.

It is worth noting that in many discussions of the role of monetary policy in safeguarding financial stability, the question is framed as asking whether monetary policy should proactively "lean against the wind" of changes in financial-market sentiment (for example, Svensson 2017). This formulation would seem to suggest that fluctuations in asset prices are an exogenous source of variation—a "wind" blowing in from outside the model, as might be associated say, with a late 1990s-style stock-market bubble driven by enthusiasm over a new technology. However, as our framework underscores, sometimes the central bank is itself the driver of movements in asset prices. In this case, the question is not whether it should lean against an external shock, but rather how aggressively a central bank should deploy a tool that itself can lead to overly compressed risk premiums.

Implications for the Conduct of Monetary Policy

Incorporating Insights from the Model into the Policy Process

How might central banks adapt their monetary-policy processes to take account explicitly of the intertemporal tradeoff we have identified? One suggestion is that policymakers should seek to develop summary measures of financial conditions that are most useful for capturing the kind of credit-bites-back risk we have highlighted. Many central banks now produce financial stability reports that track a wide variety of indicators in financial markets, which represents progress relative to the situation before the global financial crisis of 2008–2009. Most of these reports, however, stop short of making an overall judgment about the level of risk to the macroeconomy and its implications, if any, for monetary policy. For instance, the Federal Reserve's biannual Financial Stability Report offers no summary assessment of the level of risk from the areas it reviews.

This approach stands in stark contrast to the treatment of key macroeconomic factors that feature in conventional models. For example, it is hard to imagine a central bank seeking to pursue inflation targeting without a commonly agreed measure of inflation.

Such a lack of consensus as to the nature of the problem can create a situation where, as long as a large number of indicators are not flashing red, the default presumption is that monetary policymakers can simply ignore credit-bites-back effects when they go about setting their target for short-term rates. Such a

default setting may be especially problematic when, as argued by Fontanier (2022), extrapolative behavior on the part of market participants implies that the right time to begin leaning against financial imbalances is relatively early in the cycle—not when these imbalances have reached a critical level and when inadvertently popping a bubble may do considerable harm.

A related challenge is to integrate the analysis of financial risk more fully into monetary policy decision-making. The Federal Reserve currently does deep dives on financial risk four times a year and publishes much of the work in two financial stability reports. The Fed should consider discussing these risks and their implication for policy at every meeting, much as they currently do with inflation, the other major source of tradeoff they face in stabilizing real activity. After all, nobody thinks that the right way to deal with the risk of accelerating inflation is to have a default presumption that it is not a problem until the situation is indisputably critical. Careful ongoing monitoring and a willingness to take early action if needed are core to the policy process for dealing with inflation. The intertemporal tradeoff associated with credit-bites-back risk should be managed analogously.

Ultimately, these changes to the policy process should be reflected in how central banks communicate with the public and the elected representatives to whom they are accountable. For example, the Fed's annual "Statement on Longer-Run Goals and Monetary Policy Strategy" mentions the importance of financial stability as a precursor for achieving its other objectives. This framing could be adjusted to recognize that threats to these objectives can come not just from exogenous developments in financial markets, but also from the Fed's own aggressive attempts to support the economy.[2]

Relatedly, in its annual monetary policy reports to Congress, the Federal Reserve shows five interest-rate rules that are used as points of reference in policy deliberations. None of these rules take account of financial conditions. The Fed may wish to experiment with alternatives that make different judgments about how to weigh the circumstances of the moment against potential constraints on future policy.

Admittedly, the current state of research does not provide decisive guidance on how best to measure credit-bites-back risk. Thus, moving in these directions poses challenges, but our view is that having even an imperfect measure of risk, taken into account in a disciplined way, is better than ignoring the potential tradeoff. Confronting these issues head on and talking publicly about them might also spur Congress to take steps to improve the macroprudential tools that are available to regulators. Any progress on that front would also be highly desirable in its own right.

[2] In fact, since May 2019 the Fed's financial stability report has included the results from a survey it conducts of its private sector contacts regarding the near-term risks to the economy. Participants routinely cite risks emanating from monetary policy as a major source of concern. For instance, in November 2022 it was deemed to be one of the top two short-term risks to the economy.

Exogenous and Endogenous Determinants of the Neutral Real Rate of Interest

A central concept in the conduct of monetary policy is the neutral real rate of interest, often referred to as r^*, which is level of the short-term real interest rate at which output equals potential and policy is neither inflationary nor deflationary. A large body of research has found that r^* declined significantly for the US economy, as well as in several other advanced economies, in the years leading up to the onset of the COVID-19 pandemic in 2020 (for a summary, a useful starting point is Holston, Laubach, and Williams 2017). Common explanations for this decline focus on *exogenous* demographic and technological factors at the global level, such as increased savings by an aging population, a slowdown in trend productivity growth, and increased income inequality (Straub 2019).

More recently, several papers have argued that part of the decline in the neutral rate of interest r^* could instead be *endogenously* related to the prior conduct of monetary policy by means of a hysteresis effect whereby low interest rates beget the need for continued low rates in the future. One mechanism that generates such an effect works through durable goods—for example, if low rates today lead consumers to buy a lot of new cars, there will be less demand for cars going forward, and the policy rate will have to be lower (all else equal) to sustain enough aggregate demand to keep the economy at full employment (McKay and Wieland 2021). Other mechanisms can have similar consequences. For example, a period of low rates encourages mortgage borrowers to refinance, which is stimulative, but which exhausts the pool of future refinancers and hence weakens the power of this channel going forward (Berger et al. 2021; see also Greenwald 2018; Wong 2021; Beraja et al. 2019). In a similar vein, easy monetary policy can cause households to become more highly indebted, which in turn makes further stimulus less effective (Mian, Straub, and Sufi 2021).[3]

Our model offers another reason why there can be history-dependence of this sort in r^*: easy monetary policy creates a boom in asset prices, but then effectively corners policymakers into keeping policy easy for fear of creating an asset-price reversal that damages the economy.

The distinction between the exogenous/demographic/technological and the endogenous/history-dependent accounts of the neutral rate of interest r^* is of practical importance for several reasons. First, if the decline in r^* is driven by outside factors, the job of the central bank is effectively to come up with its best empirical estimate of the current (exogenous) value of r^* and then to set policy rates accordingly. By contrast, if the decline in r^* is at least partially endogenous, there is a looking-in-the-mirror problem: simply knowing that it will take a low policy rate today to maintain full employment is insufficient for making good decisions over time, because this observation muddles together exogenous factors and the history

[3] In a related vein, Acharya and Rajan (2022) and Acharya et al. (2022) emphasize a potentially history-dependent impact of the increases in bank reserves driven by quantitative easing. They observe that as reserves grow, intermediaries create additional short-term deposits and expand credit lines to match the increase in reserves. They argue that the presence of these claims can lock the central bank into needing to keep reserves high in order for the intermediaries to be able to honor these claims.

of past policy choices. In addition, it ignores the likelihood that low interest rates today may have repercussions for the future monetary policy opportunity set.

Second, the exogenous/demographic/technological view suggests that movements in the neutral rate of interest r^* are likely to be highly persistent, given that the underlying driving factors themselves are so slow-moving. Such a view seems to have informed the Fed's framework review of August 2020, which unequivocally endorsed the proposition that r^* would continue to remain low for the foreseeable future, and which adopted a "lower for longer" philosophy—one that arguably proved problematic when inflation began to rise sharply in the following year.[4] By contrast, an endogenous/history-dependent interpretation of the history of r^* would have presumably provided less confidence as to its stability over the coming years.

International Considerations

Our discussion has thus far taken a largely closed-economy perspective. But the observation that monetary policy works by influencing risk premiums also has important international implications. In influential works, Rey (2013) and Miranda-Agrippino and Rey (2020) argue that if monetary-policy-induced changes in risk premiums are highly correlated across countries—as one might expect if the arbitrageurs who police these risk premiums are global financial players—then individual central banks around the world will have less policy independence than is normally envisioned in flexible-exchange-rate, open-economy macro models.

Table 1 illustrates this point, focusing on data from the period January 1998 to December 2021. The left column of the table shows the correlation of one-month changes in one-year yields—a natural proxy for the expected short-term path of monetary policy—between US government bonds and those from six other advanced economies: Australia, Canada, Switzerland, Germany, Great Britain, and Japan. The right column repeats the exercise for ten-year yields, which one can think of as capturing both the expected path of monetary policy, as well as a term premium. As can be seen, in all cases, the correlation of changes in long-term yields is higher than the correlation of changes in short-term yields. In several cases, most notably Australia, Germany, and Great Britain, this differential is strikingly large; for example, the correlation of changes in Australian one-year yields with changes in US one-year yields is 0.42, while for ten-year yields the corresponding correlation is 0.73.

This pattern suggests that term premiums across countries are more tightly correlated than short-term policy rates, which underscores the point raised by Rey (2013) and Miranda-Agrippino and Rey (2020): even if one country's central bank attempts to set its monetary policy in a way that is independent of that in other

[4]In a speech accompanying the revised 2020 Statement on Longer-Run Goals and Monetary Policy Strategy, Chair Jerome Powell (2020) said: "This decline in assessments of the neutral federal funds rate has profound implications for monetary policy. . . . [G]oing forward, employment can run at or above real-time estimates of its maximum level without causing concern, unless accompanied by signs of unwanted increases in inflation or the emergence of other risks that could impede the attainment of our goals."

Table 1

Correlations between One-Month Changes in One-Year and Ten-Year US and Advanced Economy Government Bond Yields

Area (currency)	*Correlation with one-year US Treasury yields*	*Correlation with ten-year US Treasury yields*
Australia (dollar)	0.42	0.73
Canada (dollar)	0.71	0.84
Switzerland (franc)	0.43	0.59
European Union (euro)	0.53	0.73
Great Britain (pound)	0.56	0.77
Japan (yen)	0.18	0.33

Source: For data sources, see Kashyap and Stein (2023).
Notes: The left column shows the correlation of one-month changes in one-year yields between US government bonds and those from, respectively, Australia, Canada, Switzerland, Germany, Great Britain, and Japan. The right column repeats the exercise for ten-year yields. The sample period runs from January 1998 to December 2021.

countries, it may not fully succeed in doing so, particularly if what ultimately matters for economic activity are risk-premium-inclusive financial conditions such as longer-term rates. Moreover, if one believes that the US Federal Reserve has a preeminent role in determining these risk premiums due to the dominant role of the US dollar in international finance, then this mechanism has the potential to increase significantly the Fed's influence over other economies.

Conclusions

Our analysis is built on two well-documented findings: (1) monetary policy operates in significant part by influencing financial-market sentiment; and (2) these sentiment shifts are prone to reversals, which can impair the credit-supply mechanism and ultimately damage the real economy. Taking account of these effects has the potential to overturn some basic presumptions about how monetary policy should be conducted. Perhaps most importantly, the risk of reversals means that optimal policy no longer always completely offsets even pure negative demand shocks. Instead, policy may in some cases need to trade off the benefits of supporting the economy now against the possibility that an unwinding of financial-market sentiment could lead to worse outcomes in the future.

The broad analytics of this tradeoff are relatively straightforward, but the practical implications are not. Addressing the tradeoff raises serious measurement challenges with respect to gauging the credit-bites-back risk. It will also require standard central-bank operating practices and communication policies to be adapted in a variety of ways. We have highlighted a number of areas where further research along these lines would be especially valuable and look forward to seeing this work develop.

■ *We are grateful to Livia Amato and Zhi Zhang for outstanding research assistance, and to Tobias Adrian, Ben Bernanke, Michael Blank, Sam Hanson, Raghu Rajan, Alp Simsek, Ludwig Straub, Amir Sufi, Adi Sunderam, and Silvana Tenreyro for helpful comments on an earlier draft.*

References

Acharya, Viral V., Rahul S. Chauhan, Raghuram G. Rajan, and Sascha Steffen. 2022. "Liquidity Dependence: Why Shrinking Central Bank Balance Sheets Is an Uphill Task." Paper presented at the Federal Reserve Bank of Kansas City Jackson Hole Symposium, Jackson Hole, WY, August 27.

Acharya, Viral V., and Hassan Naqvi. 2019. "On Reaching for Yield and the Coexistence of Bubbles and Negative Bubbles." *Journal of Financial Intermediation* 38: 1–10.

Acharya, Viral V., and Raghuram G. Rajan. 2022. "Liquidity, Liquidity Everywhere, Not a Drop to Use: Why Flooding Banks with Central Bank Reserves May Not Expand Liquidity." NBER Working Paper 29680.

Adrian, Tobias, Nina Boyarchenko, and Domenico Giannone. 2019. "Vulnerable Growth." *American Economic Review* 109 (4): 1263–89.

Adrian, Tobias, and Fernando Duarte. 2020. "Financial Vulnerability and Monetary Policy." Federal Reserve Bank of New York Staff Report 804.

Adrian, Tobias, Federico Grinberg, Nellie Liang, Sheheryar Malik, and Jie Yu. 2022. "The Term Structure of Growth-at-Risk." *American Economic Journal: Macroeconomics* 14 (3): 283–323.

Anderson, Nicola, and John Sleath. 1999. "New estimates of the UK real and nominal yield curves." *B*ank of England Quarterly Bulletin 1999 Q4.

Anderson, Nicola, and John Sleath. 2001. "New estimates of the UK real and nominal yield curves." Bank of England Working Paper 126.

Bank of Canada. 2022. "Yield curves for Zero-Coupon Bonds." https://www.bankofcanada.ca/rates/interest-rates/bond-yield-curves (accessed December 10, 2022).

Bank of England. 2022. "Yield curves." https://www.bankofengland.co.uk/statistics/yield-curves (accessed December 10, 2022).

Bauer, Michael D., and Eric T. Swanson. 2022. "A Reassessment of Monetary Policy Surprises and High-Frequency Identification." In *NBER Macroeconomics Annual*, Vol. 37, edited by Martin Eichenbaum, Erik Hurst, and Valerie A. Ramey. Chicago: University of Chicago Press.

Beraja, Martin, Andreas Fuster, Erik Hurst, and Joseph Vavra. 2019. "Regional Heterogeneity and the Refinancing Channel of Monetary Policy." *Quarterly Journal of Economics*, 134 (1): 109–83.

Berger, David, Konstantin Milbradt, Fabrice Tourre, and Joseph Vavra. 2021. "Mortgage Prepayment and Path-Dependent Effects of Monetary Policy." *American Economic Review* 111 (9): 2829–78.

Bernanke, Ben S. 2014. "A Conversation: The Fed Yesterday, Today, and Tomorrow." The Brookings Institution, January 16. https://www.brookings.edu/wp-content/uploads/2014/01/20140116_bernanke_remarks_transcript.pdf.

Bernanke, Ben S. 2015. "Federal Reserve Policy in an International Context." Paper presented at the 16th Jacques Polak Annual Research Conference Hosted by the International Monetary Fund. Washington, DC, November 5.

Bernanke, Ben S. 2020. "The New Tools of Monetary Policy." *American Economic Review* 110 (4): 943–83.

Bernanke Ben S., Mark Gertler, and Simon Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics*, Vol. 1C, edited by John B. Taylor and Michael Woodford, 1341–93. Amsterdam: North-Holland.

Bernanke, Ben S., and Kenneth N. Kuttner. 2005. "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *Journal of Finance* 60 (3): 1221–57.

Bloomberg L.P. n.d. "Japan Government Bond 6-month Yield [GTJPY6M Govt]", provided via Bloomberg Terminal Data Subscription (accessed December 10, 2022).

Bloomberg L.P. n.d. "Japan Government Bond 1-Year Yield [GTJPY1Y Govt]", provided via Bloomberg Terminal Data Subscription (accessed December 10, 2022).

Board of Governors of the Federal Reserve System (US). 2022a. "Nominal Yield Curve." https://www.federalreserve.gov/data/nominal-yield-curve.htm (accessed December 10, 2022).

Board of Governors of the Federal Reserve System (US). 2022b. "Term Premium on a 10 Year Zero Coupon Bond [THREEFYTP10]", retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/THREEFYTP10 (accessed December 11, 2022).

Board of Governors of the Federal Reserve System (US). 2022c. "Net Percentage of Domestic Banks Tightening Standards for Commercial and Industrial Loans to Large and Middle-Market Firms [DRTSCILM]", retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/DRTSCILM (accessed December 11, 2022).

Board of Governors of the Federal Reserve System (US). 2022d. "Senior Loan Officer Opinion Survey on Bank Lending Practices." https://www.federalreserve.gov/data/sloos.htm.

Bolder, David, Adam Metzler, and Grahame Johnson. 2004. "An Empirical Analysis of the Canadian Term Structure of Zero-Coupon Interest Rates." Bank of Canada Staff Working Paper 2004-48.Borio, Claudio, and Haibin Zhu. 2012. "Capital Regulation, Risk-Taking and Monetary Policy: A Missing Link in the Transmission Mechanism?" *Journal of Financial Stability* 8 (4): 236–51.

Boyarchenko, Nina, Anna Kovner, and Or Schachar. 2020. "It's What You Say and What You Buy: A Holistic Evaluation of the Corporate Credit Facilities." Federal Reserve Bank of New York Staff Report 935.

Boyarchenko, Nina, Giovanni Favara, and Moritz Schularick. 2022. "Financial Stability Considerations for Monetary Policy: Empirical Evidence and Challenges." Federal Reserve Bank of New York Staff Report 1003.

Caballero, Ricardo J., and Alp Simsek. 2020. "Prudential Monetary Policy." NBER Working Paper 25977.

Caballero, Ricardo J., and Alp Simsek. 2022. "A Monetary Policy Asset Pricing Model." NBER Working Paper 30132.

Campbell, John Y., and Roman Sigalov. 2022. "Portfolio Choice with Sustainable Spending: A Model of Reaching for Yield." *Journal of Financial Economics* 143 (1): 188–206.

Carpenter, Seth B., Ethan S. Harris, Peter Hooper, Anil K. Kashyap, and Kenneth D. West. 2022. "Some Benefits and Risks of a Hot Economy." US Monetary Policy Forum 2022 Report.

Chodorow-Reich, Gabriel. 2014. "Effects of Unconventional Monetary Policy on Financial Institutions." *Brookings Papers on Economic Activity* 44 (1): 155–204.

Cieslak, Anna, and Hao Pang. 2021. "Common Shocks in Stocks and Bonds." *Journal of Financial Economics* 142 (2): 880–904.

Clarida, Richard H., Burcu Duygan-Bump, and Chiara Scotti. 2021. "The COVID-19 Crisis and the Federal Reserve's Policy Response." FEDS Working Paper 2021-35.

D'Amico, Stefania, Vamsi Kurakula, and Stephen M. Lee. 2020. "Impacts of the Fed Corporate Credit Facilities through the Lenses of ETFs and CDX." Federal Reserve Bank of Chicago Working Paper WP-2020-14.

Dell'Ariccia, Giovanni, Luc Laeven, and Gustavo A. Suarez. 2017. "Bank Leverage and Monetary Policy's Risk-Taking Channel: Evidence from the United States." *Journal of Finance* 72 (2): 613–54.

Deutsche Bundesbank Eurosystem. 2022. "Interest rates and yields." https://www.bundesbank.de/dynamic/action/en/statistics/time-series-databases/time-series-databases/759784/759784?listId=www_skms_it03a (accessed December 10, 2022).

Di Maggio, Marco, and Marcin Kacperczyk. 2017. "The Unintended Consequences of the Zero Lower Bound Policy." *Journal of Financial Economics* 123 (1): 59–80.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl. 2017. "The Deposits Channel of Monetary Policy." *Quarterly Journal of Economics* 132 (4): 1819–76.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl. 2018. "A Model of Monetary Policy and Risk Premia." *Journal of Finance* 73 (1): 317–73.

Finlay, Richard, and Mark Chambers. 2008. "A Term Structure Decomposition of the Australian Yield Curve." Reserve Bank of Australia Research Discussion Paper 2008–09.

Finlay, Richard, and David Olivan. 2012. "Extracting Information from Financial Market Instruments." Reserve Bank of Australia Bulletin March.

Fontanier, Paul. 2022. "Optimal Policy for Behavioral Financial Crises." Working Paper. https://paulfontanier.github.io/papers/macropru_behav_nov22.pdf.

Galí, Jordi. 2008. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework: An Introduction to the New Keynesian Framework*. Princeton, NJ: Princeton University

Press.
Galí, Jordi. 2018. "The State of New Keynesian Economics: A Partial Assessment." *Journal of Economic Perspectives* 32 (3): 87–112.
Gertler, Mark, and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7 (1): 44–76.
Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102 (4): 1692–720.
Gourinchas, Pierre-Olivier, Walker D. Ray, and Dimitri Vayanos. 2022. "A Preferred-Habitat Model of Term Premia, Exchange Rates, and Monetary Policy Spillovers." NBER Working Paper 29875.
Greenwald, Daniel. 2018. "The Mortgage Credit Channel of Macroeconomic Transmission." MIT Sloan Research Paper 5184-16.
Greenwood, Robin, and Samuel G. Hanson. 2013. "Issuer Quality and Corporate Bond Returns." *Review of Financial Studies* 26 (6): 1483–525.
Greenwood, Robin, Samuel G. Hanson, Andrei Shleifer, and Jakob A. Sørensen. 2022. "Predictable Financial Crises." *Journal of Finance* (77) 2: 863–921.
Greenwood, Robin, Samuel G. Hanson, Jeremy C. Stein, and Adi Sunderam. Forthcoming. "A Quantity-Driven Theory of Term Premia and Exchange Rates." *Quarterly Journal of Economics***.**
Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright. 2007. "The U.S. Treasury yield curve: 1961 to the present." *Journal of Monetary Economics* 54 (8): 2291–2304.
Hanson, Samuel G., and Jeremy C. Stein. 2015. "Monetary Policy and Long-Term Real Rates." *Journal of Financial Economics* 115 (3): 429–48.
Hanson, Samuel G., David O. Lucca, and Jonathan H. Wright. 2021. "Rate-Amplifying Demand and the Excess Sensitivity of Long-Term Rates." *Quarterly Journal of Economics* 136 (3): 1719–81.
Holston, Kathryn, Thomas Laubach, and John C. Williams. 2017. "Measuring the Natural Rate of Interest: International Trends and Determinants." *Journal of International Economics* 108 (S1): S59–75.
Ioannidou, Vasso, Steven Ongena, and José-Luis Peydró. 2015. "Monetary Policy, Risk-Taking, and Pricing: Evidence from a Quasi-natural Experiment." *Review of Finance* 19 (1): 95–144.
Jiménez, Gabriel, Steven Ongena, José-Luis Peydró, and Jesús Saurina. 2014. "Hazardous Times for Monetary Policy: What Do Twenty-Three Million Bank Loans Say about the Effects of Monetary Policy on Credit Risk-Taking?" *Econometrica* 82 (2): 463–505.
Jordà, Òscar, Moritz Schularick, and Alan M. Taylor. 2013. "When Credit Bites Back: Leverage, Business Cycles, and Crises." *Journal of Money, Credit and Banking* 45 (S2): S3–28.
Kashyap, Anil, K., and Jeremy C. Stein. 2000. "What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?" *American Economic Review* 90 (3): 407–28.
Kashyap, Anil K., and Jeremy C. Stein. 2023. "Replication data for: Monetary Policy When the Central Bank Shapes Financial-Market Sentiment." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E183525V1.
Kekre, Rohan, and Moritz Lenel. 2022. "Monetary Policy, Redistribution, and Risk Premia." University of Chicago Becker Friedman Institute for Economics Working Paper 2020-02.
Kim, Don H., and Jonathan H. Wright. 2005. "An Arbitrage-Free Three-Factor Term Structure Model and the Recent Behavior of Long-Term Yields and Distant-Horizon Forward Rates." FEDS Working Paper 2005-33.
Kirti, Divya. 2018. "Lending Standards and Output Growth." IMF Working Paper 2018/23.
Krishnamurthy, Arvind, and Tyler Muir. 2020. "How Credit Cycles across a Financial Crisis." Stanford Graduate School of Business Working Paper 3579.
Krishnamurthy, Arvind, and Annette Vissing-Jørgensen. 2011. "The Effects of Quantitative Easing on Interest Rates: Channels and Implications for Policy." *Brookings Papers on Economic Activity* 2011.
Lian, Chen, Yueran Ma, and Carmen Wang. 2019. "Low Interest Rates and Risk-Taking: Evidence from Individual Investment Decisions." *Review of Financial Studies* 32 (6): 2107–48.
López-Salido, David, Jeremy C. Stein, and Egon Zakrajšek. 2017. "Credit-Market Sentiment and the Business Cycle." *Quarterly Journal of Economics* 132 (3): 1373–426.
Maddaloni, Angela, and José-Luis Peydró. 2011. "Bank Risk-Taking, Securitization, Supervision, and Low Interest Rates: Evidence from the Euro-Area and the U.S. Lending Standards." *Review of Financial Studies* 24 (6): 2121–65.
McKay, Alisdair, and Johannes F. Wieland. 2021. "Lumpy Durable Consumption Demand and the Limited Ammunition of Monetary Policy." *Econometrica* 89 (6): 2717–49.
Mian, Atif, Ludwig Straub, and Amir Sufi. 2021. "Indebted Demand." *Quarterly Journal of Economics* 136 (4): 2243–307.

Mian, Atif, Amir Sufi, and Emil Verner. 2017. "Household Debt and Business Cycles Worldwide." *Quarterly Journal of Economics* 132 (4): 1755–817.

Ministry of Finance, Japan. 2022. "Interest rate." https://www.mof.go.jp/english/policy/jgbs/reference/interest_rate/index.htm (accessed December 10, 2022).

Miranda-Agrippino, Silvia, and Hélène Rey. 2020. "U.S. Monetary Policy and the Global Financial Cycle." *Review of Economic Studies* 87 (6): 2754–76.

Palazzo, Berardino, and Ram Yamarthy. 2022. "Credit Risk and the Transmission of Interest Rate Shocks." *Journal of Monetary Economics* 130: 120–36.

Paligorova, Teodora, and João A.C. Santos. 2017. "Monetary Policy and Bank Risk-Taking: Evidence from the Corporate Loan Market." *Journal of Financial Intermediation* 30: 35–49.

Peek, Joe, Eric S. Rosengren, and Geoffrey M. B. Tootell. 2016. "Does Fed Policy Reveal a Ternary Mandate?" Federal Reserve Bank of Boston Working Paper 16-11.

Powell, Jerome H. 2020. "New Economic Challenges and the Fed's Monetary Policy Review." Speech, Federal Reserve Bank of Kansas City Symposium, Jackson Hole, WY, August 27, 2020. https://www.federalreserve.gov/newsevents/speech/powell20200827a.htm.

Rajan, Raghuram G. 2005. "Has Financial Development Made the World Riskier?" Paper presented at the Federal Reserve Bank of Kansas City Economic Policy Symposium, Jackson Hole, WY.

Razzak, Ramiz Aziz. 2022. "Leaning against Credit? Evidence that the Fed Incorporates Financial Stability Considerations into Monetary Policy." Bachelor's thesis. Harvard University.

Reserve Bank of Australia. 2022a. "Zero-coupon interest rates - 1992 to 2008 - F17." https://www.rba.gov.au/statistics/historical-data.html (accessed December 10, 2022).

Reserve Bank of Australia. 2022b. "Zero-coupon interest rates - 2009 to current - F17." https://www.rba.gov.au/statistics/tables/#interest-rates (accessed December 10, 2022).

Rey, Hélène. 2013. "Dilemma Not Trilemma: The Global Financial Cycle and Monetary Policy Independence." Paper presented at the Federal Reserve Bank of Kansas City Economic Policy Symposium, Jackson Hole, WY, August 24.

Schularick, Moritz, and Alan M. Taylor. 2012. "Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crises, 1870–2008." *American Economic Review* 102 (2): 1029–61.

Shiller, Robert J. 2016. Irrational Exuberance. Revised and expanded 3rd ed. Princeton: Princeton University Press.

Shiller, Robert J. 2022. "Data on Cyclically Adjusted Price Earnings Ratio P/E10 [CAPE]." http://www.econ.yale.edu/~shiller/data.htm (accessed December 11, 2022).

Stein, Jeremy C. 2021. "Can Policy Tame the Credit Cycle?" *IMF Economic Review* 69 (1): 5–22.

Straub, Ludwig. 2019. "Consumption, Savings, and the Distribution of Permanent Income." https://scholar.harvard.edu/files/straub/files/cons_ineq_rates.pdf.

Sufi, Amir, and Alan M. Taylor. 2021. "Financial Crises: A Survey." University of Chicago Becker Friedman Institute Working Paper 2021-97.

Svensson, Lars E.O. 2017. "Cost-Benefit Analysis of Leaning against the Wind." *Journal of Monetary Economics* 90: 193–213.

Swiss National Bank. 2022a. "Interest rates. Yields on bond issues" https://data.snb.ch/en/topics/ziredev/cube/rendoblid (accessed December 10, 2022).

Swiss National Bank. 2022b. Quarterly Bulletin 2/2002.

Wong, Arlene. 2021. "Refinancing and the Transmission of Monetary Policy to Consumption." https://static1.squarespace.com/static/576576adbe659449f97e0d35/t/60770bfec442b6095624f676/1618414592249/Paper_AERrevision2.pdf.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton, NJ: Princeton University Press.

Journal of Economic Perspectives—Volume 37, Number 1—Winter 2023—Pages 77–100

Risk Appetite and the Risk-Taking Channel of Monetary Policy

Michael D. Bauer, Ben S. Bernanke, and Eric Milstein

How does monetary policy affect the economy? Traditional macroeconomic models posit that monetary policy works primarily through three *neoclassical channels*: cost-of-capital effects, wealth effects, and exchange-rate effects. To illustrate these channels, consider a situation where the central bank raises interest rates in order to prevent the economy from overheating. First, the increase in the cost of capital will dissuade capital investments by firms and purchases of houses and durables by consumers. Second, higher rates will reduce the present value of various assets and the resulting wealth effects will lower aggregate spending. Third, higher rates will strengthen the domestic currency, depressing net exports. In addition, a more modern view recognizes the importance of frictions in financial markets, so that monetary policy may also affect economic activity via so-called *credit channels*. For example, tighter policy reduces both the net worth and the cash flow of firms, and these balance-sheet effects make it more expensive for them to obtain external financing, depressing investment.[1]

These standard channels are important, but they typically place little or no weight on changes in risk perceptions and risk attitudes. This omission is potentially important, because fluctuations in people's willingness to take risks naturally affect their economic decisions. Considerable evidence suggests that the propensities of

■ *Michael Bauer is Professor of Economics, Universität Hamburg, Hamburg, Germany. Ben Bernanke is Distinguished Senior Fellow, Brookings Institution, Washington, DC. Eric Milstein is a PhD student, University of Chicago, Chicago, Illinois. Their email addresses are michael.bauer@uni-hamburg.de, bbernanke@brookings.edu, and emilstein@uchicago.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.1.77.

[1] For further discussion of the channels of monetary transmission, see Boivin, Kiley, and Mishkin (2010). On credit channels, see Bernanke and Gertler (1995) and Bernanke, Gertler, and Gilchrist (1999).

lenders, borrowers, investors, and other economic actors to take risks do indeed vary over time. Moreover, the willingness to take risks is likely influenced by the stance of monetary policy, with easier policy associated with a greater appetite for risk and tighter policy linked to reduced risk appetite. The tendency of monetary policy to affect macroeconomic conditions by changing risk-taking and risk premia has been dubbed the *risk-taking channel* of monetary transmission (Borio and Zhu 2012). Rather than describing a single, specific mechanism, this channel can include a variety of mechanisms operating via financial intermediaries, institutional investors, or the behavior of households.

In this article, we discuss the role of shifts in risk appetite in the transmission of monetary policy to financial markets and the macroeconomy. Our main focus is to review and extend the empirical evidence on the effects of monetary policy on risk appetite in financial markets, the first stage of the risk-taking channel. To identify these effects, we consider high-frequency changes in financial markets, following recent empirical literature in monetary economics (Nakamura and Steinsson 2018a, b). Specifically, we use event studies of the effects of announcements by the Federal Open Market Committee (FOMC) on risky asset prices. We use financial market data around FOMC announcements to measure the unexpected component of monetary policy actions. The event studies generally show that these "monetary policy surprises" have substantial effects on the prices of various risky assets. Consistent with the risk-taking channel, unexpected policy easing leads to "risk on" changes in financial markets, including higher stock returns, lower stock market volatility, tighter credit spreads, and a weaker dollar. Similarly, unexpected tightening leads to "risk off" changes and the opposite movements in risky asset prices.

An important question is whether these estimated effects, many of which have been documented previously, are due to changes in the overall risk appetite of investors or arise from other sources, such as changes in fundamentals or the perceived riskiness of specific assets. To address this question, we develop a new index of risk appetite in financial markets based on the common component of various risk indicators from equity, fixed income, credit, and foreign exchange markets. Our working assumption, motivated by standard asset-pricing theory, is that common movements in risk premia and risky asset prices across all of these markets are due primarily to changes in the overall level of risk appetite. Using our new index, we study changes in risk appetite around FOMC announcements. We find that monetary policy actions appear to have strong and persistent effects on risk appetite, which drive a substantial component of the transmission of monetary policy to financial markets.

Although we do not provide direct evidence of the macroeconomic effects of the risk-taking channel, our results are consistent with a quantitatively important role for this channel in the transmission of monetary policy to the real economy. Changes in risk appetite and risk premia are key determinants of asset prices, wealth, collateral values, and credit costs, which in turn affect financing and spending decisions through a variety of conduits. We would also expect risk appetite in financial

markets to be highly correlated with the willingness of banks, firms, and households to take risks in their lending, investment, and borrowing decisions. However, more empirical research is needed to quantify the macroeconomic importance of the risk-taking channel, as well as to understand its implications for optimal policy and financial stability.

Time Variation in Risk Appetite

A key premise of the risk-taking channel is that the risk appetite of investors and other economic agents changes over time. In this section, we discuss why such changes might occur and why monetary policy might be a source of such changes.

The return to any financial asset includes a *risk premium*, that is, the extra compensation that investors receive for bearing the risk of that asset. The risk premium of an asset can usefully be conceptualized as the product of the *price of risk* and the *quantity of risk*—the compensation investors require for each "unit" of risk in their portfolios, times the amount of undiversifiable risk inherent in each specific asset. Indeed, most standard asset pricing models lead to such an intuitive decomposition of risk premia (Cochrane 2005).

Risk appetite, the willingness of investors to bear risk, is typically defined as the inverse of the price of risk (for example, Gai and Vause 2006). The economy-wide level of risk appetite affects risk premia in all financial markets—that is, it is common to all real and financial assets. By contrast, the quantity of risk is asset-specific and depends on the distribution of the particular asset's possible future returns. In general, an asset is riskier if it tends to have high payoffs in states of the world in which investors have high levels of consumption and thus low marginal utility. Because such an asset does not hedge against the risk of bad consumption outcomes, it is less valuable (all else equal) and investors will require greater compensation (in the form of a higher risk premium) to be willing to own it. Because assets differ in their characteristic quantity of risk, risk premia will differ across assets even if the price of risk is the same for all assets.

Why Might Risk Appetite Vary?

Much evidence and casual observation suggest that investors' risk appetite varies over time. What explains that variation? The classic consumption-based asset pricing model provides some intuition: in this model, the price of risk is the product of the representative agent's variance of consumption growth and the agent's degree of risk aversion (as determined by the curvature of the agent's utility function; see Cochrane 2005, p. 17). The classic version of the model is too stylized to be useful in practice; for example, the assumption that consumption growth is the only source of risk is too restrictive. But the model is helpful because it suggests that risk appetite may vary for two broad reasons: shifts in the economic outlook and in investors' risk preferences. We consider each of these in turn.

First, risk appetite changes when the economic or financial outlook changes. For example, risk appetite is likely to improve if the economic outlook becomes more favorable—with the result, say, of raising the mean or reducing the variance of future consumption. This link is at the core of many asset pricing theories that focus on time variation in economic uncertainty and consumption risks to generate changing risk appetite, including long-run risk models (Bansal and Yaron 2004) and models with variable consumption disasters (Wachter 2013). The economic and financial outlook can also affect risk appetite indirectly through its effect on asset values and balance sheets. Because of asymmetric information and other frictions in credit markets, stronger lender and borrower balance sheets are associated with increased credit extension and more-rapid economic growth (Bernanke, Gertler, and Gilchrist 1999), which raises risk appetite. In extreme situations like the 2008–2009 financial crisis, widespread concerns about the solvency of lenders (including critical financial institutions) and borrowers (including both households and firms) can cause a sharp decline in risk appetite.

Second, risk appetite can change because of shifts in the underlying attitudes of investors towards risk, that is, because of time-varying risk aversion. Both finance practitioners and researchers have commonly observed that investors appear to alternate between bouts of optimism and pessimism, sometimes called "risk on, risk off" behavior. Such changes in sentiment are often cited as explanations of violent swings in financial markets, including the rapid shifts from inflows to outflows of capital from emerging-market economies (Chari, Stedman, and Lundblad 2020; Forbes and Warnock 2021) and the periodic "flights to safety," when many investors seek to increase their holdings of safe assets like US Treasury debt (Baele et al. 2020).

Modeling these swings in investor risk attitudes is challenging, and various approaches have been proposed in the asset pricing literature to generate time-varying risk aversion. One particularly influential strand of this literature has relied on habit formation in consumption, as in the seminal contribution of Campbell and Cochrane (1999). In habit formation models, people are assumed to become accustomed to their recent levels of consumption and thus more risk-averse to gambles that could result in current consumption falling close to or even below habitual levels. By the same token, risk aversion falls as people's expected consumption rises relative to its habitual level. Habit formation models thus imply that risk appetite is procyclical, rising during expansions (when consumption is high) and falling during recessions.

Balance sheet constraints of financial intermediaries can also lead to changes in effective risk aversion. The basic mechanism is that a decline in the aggregate level of capital of intermediaries, by increasing their leverage, brings them closer to regulatory or self-imposed risk limits and therefore reduces their willingness to take on risks. Models of "intermediary asset pricing" give a central role to such constraints in explaining changes in risk appetite and risk premia (for example, Adrian and Shin 2010; He and Krishnamurthy 2013).

Consumption-based and intermediary-based asset pricing models draw a tight connection between normally slow-moving fundamentals and risk aversion, which makes it challenging to generate the relatively frequent changes in investor risk aversion observed in some contexts. Other theories allow more flexibility, for example, by assuming that risk aversion can shift over time for reasons unrelated to fundamentals. A notable example is the "moody investor" framework of Bekaert, Engstrom, and Grenadier (2010), which allows for spontaneous changes in investor sentiment (see also Bekaert, Engstrom, and Xu 2022).

Variation over time in risk appetite has also been explained by "reach for yield"—the idea that investors target a certain return on their assets. When interest rates are low, they accept greater risks—that is, they effectively become less risk-averse—to give themselves a chance to earn their desired return, even though they also increase their risk of loss (for example, Hanson and Stein 2015; Becker and Ivashina 2015). For financial institutions, reach-for-yield behavior might be motivated by distorted regulatory incentives or by contractual obligations. For example, a financial institution that has made prior commitments to provide customers a specified return, as with a defined-benefit pension program or certain insurance contracts, may reach for yield to meet these commitments. For individual investors, the tendency to reach for yield when interest rates are low likely has a significant behavioral component, such as a strong preference for consuming only the current return to wealth rather than drawing down accumulated savings (Lian, Ma, and Wang 2019; Campbell and Sigalov 2022). Many open questions remain about reach-for-yield phenomena, including whether investors are most influenced by the current level of the nominal interest rate, the level of the real interest rate, or the current rate relative to historical norms.[2]

Monetary Policy and Risk Appetite

This discussion suggests that, from an asset-pricing perspective, monetary policy could affect risk appetite through its impacts on both the economic environment and on investors' risk preferences.[3] Easing the stance of monetary policy could increase risk appetite by improving the perceived economic and financial environment, for example, by upgrading the economic outlook, reducing economic uncertainty, or strengthening the balance sheets of borrowers and lenders. Both the reduction of economic and financial risks (for example, lower consumption variance) as well as the improved outlook (for example, higher expected consumption relative to habit) would contribute to higher risk appetite. Alternatively, in an environment in which investors reach for yield, the low interest rates associated with easy monetary policy—and a widening gap between target rates of return and

[2]Other asset pricing theories focus on the link between investor heterogeneity and changes in aggregate risk aversion, including Chan and Kogan (2002), Gârleanu and Pedersen (2011), and Kekre and Lenel (2022).

[3]The article by Kashyap and Stein in this symposium provides additional discussion of the mechanisms through which monetary policy might affect risk appetite, as well the literature on the effects of monetary policy on risky asset prices and risk premia in various financial markets.

market rates—could make investors effectively less risk-averse, relative to a situation in which policy was tighter and rates were higher.

Much of the literature on the risk-taking channel of monetary policy thus far has focused on risk-taking by financial institutions and the effects of monetary policy on intermediaries' profits, access to funding, leverage, and, ultimately, the volume and riskiness of their lending (Adrian and Shin 2010; Drechsler, Savov, and Schnabl 2018). The available evidence generally supports the existence of a risk-taking channel working through financial institutions, with monetary easing causing them to make more and riskier loans (as in Bruno and Shin 2015; Paligorova and Santos 2017). But while changes in the risk appetite of financial institutions are likely important for monetary transmission, these effects are only a subset of the risk-taking channel, broadly defined (Borio and Zhu 2012). If monetary policy has powerful effects on risk appetite and risky asset prices, then the more-traditional channels of monetary transmission, including wealth effects, changes in the cost of capital, and changes in borrower creditworthiness, are likely to be amplified as well (as in Bernanke 2007; Disyatat 2011). That observation motivates the study of the connection of monetary policy and risk appetite in general, not only in the context of financial institutions.

Monetary Policy Surprises and the Prices of Risky Assets

Previous empirical research on the risk-taking channel has documented substantial effects of monetary policy on risky asset prices and risk premia in various financial markets, including stock, bond, and credit markets. Bernanke and Kuttner (2005) found that monetary easing raises stock prices, not only by lowering the risk-free discount rate and raising expected future dividends, as in the traditional analysis, but to an important degree by reducing the risk premium that investors demand to hold stocks. Hanson and Stein (2015) documented a surprisingly large response of long-term real bond yields to changes in the policy rate and argued that this can only be explained if monetary policy affects the term premium. Hanson, Lucca, and Wright (2021) similarly argue that this excess sensitivity of long-term rates requires that changes in short-term rates move the term premium in the same direction, at least temporarily. Gertler and Karadi (2015) showed that monetary policy affects credit costs in large part through its effects on term premiums and credit spreads, rather than through changes in the safe rate of interest (see also Gilchrist, López-Salido, and Zakrajšek 2015). Bekaert, Hoerova, and Lo Duca (2013) developed proxies for the levels of risk and uncertainty perceived by investors and found that both, but especially risk, respond to changes in the stance of monetary policy. Miranda-Agrippino and Rey (2020) found that easier US monetary policy increases the return to risky assets globally.

Since efficient markets incorporate publicly available information, it is important that estimations of the effects of monetary policy on asset prices incorporate only *unanticipated* policy changes. In an important paper, Kuttner (2001) showed

how to measure unanticipated policy changes by using data from the market for federal funds futures, in which investors make bets on future values of the federal funds interest rate. By comparing the target for the funds rate announced by the FOMC after its policy meeting to the value previously expected by traders in the fed funds futures market, Kuttner estimated the surprise component of the change in the fed funds target rate. Regression of changes in asset prices over a short window around FOMC announcements on this *monetary policy surprise* yields an estimate of the impact of unanticipated policy changes on those asset prices.[4] The underlying idea is that the policy action was determined based on data available before the event window, ruling out reverse causality running from changes in asset prices to the policy action.

Kuttner's (2001) insight has been extended and a number of alternative measures of monetary policy surprises are now available. Gürkaynak, Sack, and Swanson (2005) incorporated information from various futures contracts related to future short-term interest rates, covering market expectations for interest rates beyond the current meeting and collectively spanning the expected path of future short rates out to a horizon of about one year. In addition, they used high-frequency data in order to measure monetary policy surprises over a tight window of 30 minutes around the announcement, which substantially improves the precision of the estimates relative to Kuttner's daily windows. Gürkaynak et al. showed that their monetary surprise can be divided into two parts: a "target factor" that measures news about the current target for the funds rate and is conceptually similar to Kuttner's measure, and a "path factor" that includes news about the funds rate's expected future path and thus captures the Fed's forward guidance about monetary policy. Gürkaynak et al. found that the path factor played an important role in determining long-term bond yields and other asset prices. Nakamura and Steinsson (2018b) used the same futures contracts as Gürkaynak et al., but measured the policy surprise series as the first principal component (that is, the main common factor) of the high-frequency futures rate changes around FOMC announcements; their policy surprise approximately corresponds to the average of the target and path factors. Bauer and Swanson (2022) constructed a similar monetary surprise measure but revised and extended the dates and times of FOMC announcements back to 1988.

Here we revisit and extend the evidence on how monetary policy affects individual risky asset prices. In the next section, we will consider policy effects on a measure that we believe better isolates the risk appetite of investors. The independent variable in our event study regressions is the measure of monetary surprises from Bauer and Swanson (2022). Like the other high-frequency measures

[4]The financial market reaction to FOMC announcements may also reflect non-conventional effects, including "information effects," which arise when the central bank's announcements reveal its private information about the state of the economy (Nakamura and Steinsson 2018b; Cieslak and Schrimpf 2019; Jarociński and Karadi 2020), or misperceptions about the Fed's systematic response to economic conditions (Bauer and Swanson 2022; forthcoming). See Bauer and Swanson (forthcoming) for more discussion.

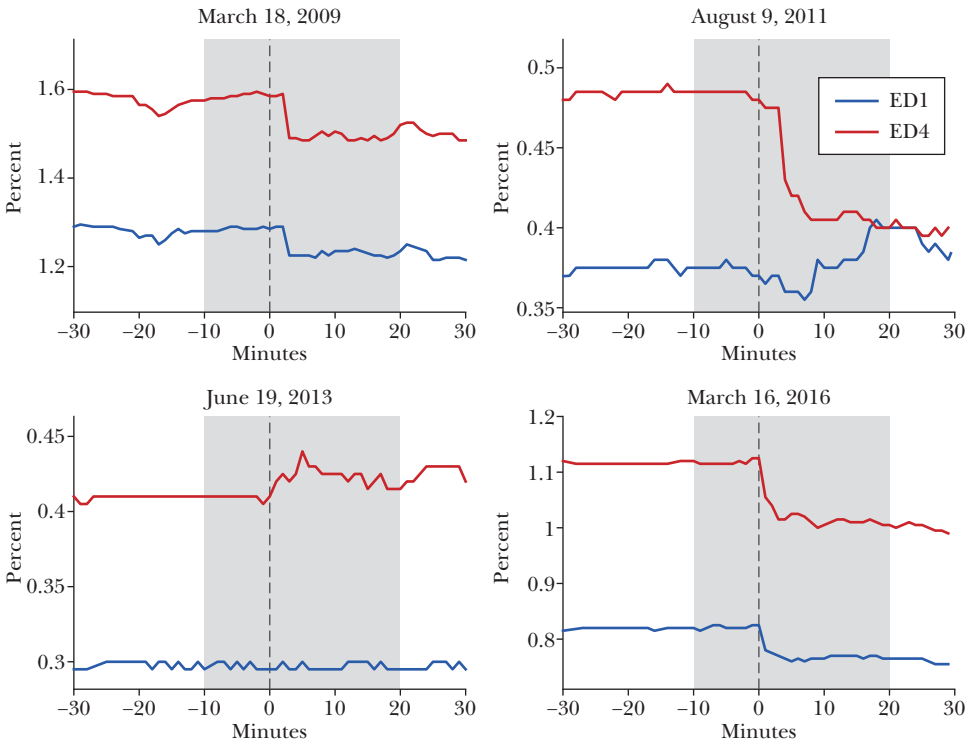
mentioned above, this measure is based on changes in interest rates over a tight intraday window around the FOMC announcement, from 10 minutes before until 20 minutes after the announcement. The calculation is based on changes in the interest rates on Eurodollar futures, which are derivative contracts with payoffs tied to the three-month London Interbank Offered Rate (LIBOR). This interest rate is an important benchmark for short-term lending in US dollars, and it is directly affected by changes in the Fed's policy rate.[5] The surprise measure is the first principal component of the changes in the first four quarterly Eurodollar futures rates (ED1 to ED4), which capture expectations of the policy rate over the current and subsequent three quarters.

Figure 1 illustrates how monetary policy surprises capture the unanticipated component of FOMC decisions. It plots the evolution of the ED1 and ED4 rates (omitting the other two Eurodollar rates used in our analysis) around four consequential FOMC announcements. Because ED1 is tied to the short rate at the end of the current quarter, it captures the market surprise about the current funds rate target decision, as well as changes in very near-term expectations. By contrast, ED4 reflects expectations for short-term rates at a horizon of about one year and therefore captures changes in more distant rate expectations, arising for example from the Fed's forward guidance and other communications.

The four plotted announcements are interesting in that none involved a change in the Fed's target of the FOMC for the federal funds rate (largely because the funds rate was already near the zero lower bound throughout most of the period), yet all triggered changes in market interest rates and policy expectations. On March 18, 2009, the Fed announced a major expansion of its first asset purchase program, commonly known as quantitative easing. The resulting decline in the market's rate expectations likely reflected the signaling effect of the dramatic new program, which was perceived as underscoring the FOMC commitment to keeping policy easier for longer (Bauer and Rudebusch 2014). On August 9, 2011, the FOMC statement included, for the first time, date-based forward guidance, as the Committee made clear its plans to avoid raising the funds rate "at least through mid-2013." This guidance substantially lowered rate expectations, causing the ED4 rate to fall by close to ten basis points. On June 19, 2013, the statement (and, later, the chair's press conference, which is not captured by the monetary surprise) raised the possibility that the Fed would soon slow ("taper") its asset purchases. Consistent with the increase in ED4 around the announcement, market participants worried that a slowing of asset purchases would be a precursor to faster rate increases than had previously been expected. The resulting volatility in bond markets became known as the "taper tantrum." Finally, on March 16, 2016, the FOMC statement signaled

[5]During normal times, three-month LIBOR is only slightly higher than the federal funds rate, but during periods of elevated financial stress, the spread between the two rates can become substantial. For example, around the March 2009 FOMC accouncement discussed below, this spread was around one percentage point. For further discussion, see, for example, Bauer, Lakdawala, and Mueller (2022).

Figure 1
Monetary Policy Surprises



Source: Tick Data.
Notes: Evolution of current-quarter (ED1) and three-quarters-ahead (ED4) Eurodollar futures rates around four important FOMC announcements. Vertical dashed lines indicate the release time of the FOMC statement. Gray-shaded areas indicate the 30-minute window used in the construction of the monetary policy surprise. The Bauer and Swanson (2022) monetary policy surprises surrounding the 2009, 2011, 2013, and 2016 meetings were –6.7, –2.0, 0.2, and –9.1 basis points, respectively.

that a tightening of policy that had been expected by markets would be deferred, resulting in a significant easing surprise.

The monetary policy surprise measure of Bauer and Swanson (2022) captures all these different types of news about monetary policy in a single number, with negative values corresponding to easing/dovish surprises, and positive numbers to tightening/hawkish surprises. It is scaled to have a one-for-one impact on ED4, that is, on one-year-ahead interest rate expectations. For example, the surprise on March 16, 2016, was –9.1 basis points, reflecting the decline in all four eurodollar futures rates in response to the FOMC announcement. For the empirical results of this paper, we rescale the Bauer-Swanson series to gauge the effects of a surprise that

leads to an increase in one-year expectations of ten basis points, a sizeable but not uncommon surprise (the standard deviation of the original policy surprises in our sample is about six basis points).[6]

Using the event-study method, we estimate the effects of monetary policy surprises on six daily variables that reflect, among other things, the risk appetite of investors: (1) the S&P 500 stock market index; (2) the S&P 500 volatility index (VIX), which measures expected stock market volatility using index option prices; (3) the spread of an index of long-term Baa-rated corporate bond yields over ten-year Treasury yields, a measure of the investment-grade credit spread; (4) a high-yield option-adjusted spread (HY OAS), which is a measure of the high-yield credit spread that adjusts for the ability of a debt issuer to call back bonds and then issue new debt if interest rates decline; (5) the spread of the three-month commercial paper rate over the federal funds rate; and (6) the trade-weighted US dollar exchange rate against advanced foreign economies. Our prior is that a surprise tightening of monetary policy, and the resulting reduction in risk appetite, should lower stock prices, increase the volatility of equities, increase the three credit spreads, and strengthen the dollar (a safe-haven currency).

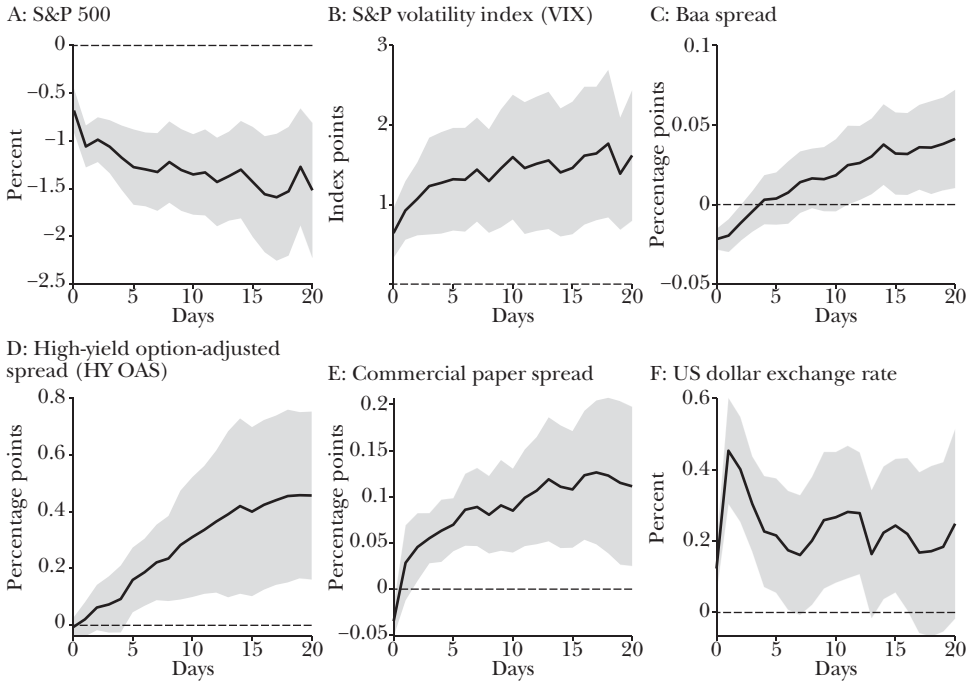
We extend previous event-study analysis of FOMC announcements and allow for both contemporaneous and lagged effects of policy surprises on asset prices. Specifically, we estimate separate regressions using different window lengths for the dependent variable: the contemporaneous asset price response on announcement days, and the cumulative responses over the subsequent 1–20 trading days. The estimated responses of each variable, together with 90-percent confidence intervals using robust standard errors, are shown in Figure 2.

The contemporaneous responses in the stock and foreign exchange markets are consistent with our priors: An unanticipated tightening of monetary policy reduces stock prices, increases stock volatility, and strengthens the dollar. By contrast, none of the three credit spreads increase immediately in response to this surprise, and the Baa and commercial paper spreads in fact significantly decline. This result, which is at odds with the theoretical channels described above, may reflect illiquidity and segmentation in corporate bond markets (Bao, Pan, and Wang 2011). If illiquidity or infrequent trading causes measured corporate yields to respond only after a delay, while Treasury yields rise immediately, we would expect to see the Baa corporate bond spread decline on impact.

Indeed, the dynamic responses in panels C–E of Figure 2 show that bond spreads go in the expected direction over time, increasing over the days following a surprise monetary tightening. Interestingly, the upward drift movement is not confined to the first few days; instead, there is an evident upward drift in spreads for several weeks after the announcement. For example, in the case of the Baa spread,

[6]In the Online Appendix, we present results for alternative measures of the monetary policy surprise, including the target and path factors of Gürkaynak et al. (2005), which separately capture news about the current target and the expected future path of the funds rate, and the composite surprise measure of Nakamura and Steinsson (2018b). The results are qualitatively similar across all surprise measures.

Figure 2
Effects of a Surprise Monetary Tightening



Source: For data sources and details of the calculations, see the Online Appendix.
Notes: Estimated response of asset prices to Bauer and Swanson (2022) monetary policy surprises, scaled to a ten basis point surprise, on the day of the FOMC announcement (day 0) and cumulative responses over the subsequent (1–20) trading days. S&P 500 and the US dollar exchange rate are measured as log changes (that is, returns), VIX as changes in index points, and credit spreads as changes in percentage points. The sample contains all FOMC announcements from January 1988 to December 2019. The sample for the VIX starts in January 1990, the high-yield option-adjusted spread (HY OAS) in January 1997, and the commercial paper (CP) spread in April 1997. Shaded areas correspond to 90 percent confidence intervals based on Huber-White heteroskedasticity-robust standard errors.

the initially negative response turns positive after four days and becomes statistically significant after twelve days. Again, lack of liquidity and transparency in corporate bond markets may help to explain this result, although the duration of the effect remains puzzling. There appears also to be some drift in the responses of S&P 500 and the VIX volatility index, although in those cases the drift is less pronounced.[7]

[7] The drift of the Baa spread, high-yield spread, and commercial paper spread are all statistically significant, in that the t-statistics for the difference between the 20-day responses and the FOMC-announcement-day

The magnitudes of the effects of monetary surprises on our chosen variables seem reasonably large. For example, after ten days, the ten basis point surprise is estimated to lower stock prices by 1.4 percent, raise the VIX volatility by 1.6 index points, increase the Baa, high-yield, and commercial paper spreads by 0.02 percentage points, 0.31 percentage points, and 0.08 percentage points, respectively, and strengthen the dollar by 0.3 percent.

These findings illustrate that, consistent with the economics of the risk-taking channel, a surprise tightening of monetary policy depresses the prices of selected risky assets. The effects appear quantitatively large and persistent. Somewhat surprisingly, credit spreads show considerable drift, rising steadily for several weeks after a hawkish monetary surprise. While these results are suggestive, they are qualified by the fact that these risk indicators will naturally depend on other factors besides risk appetite. For example, a surprise monetary tightening presumably lowers stock prices not only by reducing risk appetite, but also by lowering expected future dividends and raising the rate at which these dividends are discounted. Similarly, a surprise monetary tightening presumably causes a deterioration of the economic outlook and thus higher expected rates of corporate default, which would also contribute to higher corporate bond spreads. To get a stronger test of the risk-taking channel, we need a cleaner measure of investors' risk appetite.

Changes in Risk Appetite around Monetary Policy Announcements

To isolate the effects of the risk-taking channel, we construct a new index of financial risk appetite. With this index, we can then use our event study approach to look more directly at how monetary surprises affect risk appetite.

Quite a few indexes of financial risk and financial conditions already exist (for overviews, see Coudert and Gex 2008; Datta et al. 2017). To cite a few prominent examples: the excess bond premium of Gilchrist and Zakrajšek (2012) is an estimate of the overall risk premium in corporate bond spreads; the Federal Reserve Board makes use of a "global risk-on/risk-off index" based on the average of daily returns of 15 risky assets (Datta et al. 2017); the Federal Reserve Bank of Chicago's National Financial Conditions Index (NFCI), based on about 100 financial indicators and described in Brave and Butters (2011), has a so-called risk sub-index that includes the most risk-sensitive indicators; Miranda-Agrippino and Rey (2020), in their study of the global financial cycle, applied a dynamic factor model to extract a single factor from 858 monthly series of risky asset prices from around the world; and Bekaert, Engstrom, and Xu (2022) constructed a model of stock and bond returns,

responses are 3.6, 2.8, and 3.1 respectively. The analogous t-statistics for the S&P 500, the VIX volatility index, and the US dollar exchange rate are lower, equal to –1.8, 2.1, and 0.7 respectively.

which they combined with data on corporate cash flows and macroeconomic developments to estimate daily measures of risk aversion and uncertainty.[8]

With many empirical measures of risk-taking and risk aversion already available, why construct a new one? We had several motivations. First, our event study of FOMC announcements requires a measure of risk appetite at a daily (or higher) frequency. The need for daily data also dictated our use of financial variables shown in Figure 2 rather than alternative measures of risk such as capital outflows, credit growth, or leverage that are available only at lower frequencies. Second, our emphasis on measuring the short-run effects of monetary policy announcements suggested a risk index focused on daily *changes* in risk appetite, as opposed to the common approach of measuring the *level* of risk-taking. Third, as our monetary policy surprise data begin in 1988, we needed an index of risk appetite that covers a longer period than most. Fourth, recognizing that factors other than risk appetite can affect the returns to risky assets, we sought to measure risk appetite based on a sufficiently large number of risk-sensitive indicators. Finally, we wanted our measure to be transparent, simple, and easy to replicate. We are not aware of an existing index of risk appetite that meets all these conditions.

Our risk appetite index is based on 14 risk-sensitive financial indicators, listed in Table 1. All the indicators used are available at a daily frequency, with start dates listed in the table. Our indicators, which span a range of key markets, include two equity indices (measured in daily log-differences), four market-based measures of volatility in stock and bond returns (daily changes in index points), six private credit spreads (daily changes in percentage points), and two exchange rates (daily log-differences). We include among the measures of volatility the equity variance risk premium estimate of Bekaert and Hoerova (2014), which those authors find to be a good indicator of risk aversion. Exchange rates are included to capture "safe haven effects" of international investors moving to US dollar investments in times of financial stress. Of the 14 variables included in the index, eleven are available daily back to 1997, eight are available from 1990, and six from at least 1988.

All of the variables listed in Table 1 are widely viewed as being sensitive to changes in risk appetite.[9] At the same time, these variables represent different asset classes and are determined by diverse factors, including both fundamentals and risk perceptions. As discussed above, risk appetite is a common driver of all risk premia in the economy. Thus, it is reasonable to assume that the *comovement* in these series is mainly driven by changes in risk appetite. Based on this assumption, our index of risk appetite is constructed as their common component—specifically, the first principal component of the 14 series, which is the linear combination of the variables that explains the greatest share of the variance for the data as a whole. The Online

[8] In addition, various financial conditions indexes, including those maintained by Bloomberg and Goldman Sachs, aim to measure the degree to which financial conditions support economic activity and thus reflect factors in addition to risk-taking, such as the safe rate of return and market liquidity.

[9] Six of the 14 component variables appeared earlier in Figure 2. Responses of the other eight variables to monetary policy surprises are shown in the Online Appendix.

Table 1
Components of the Daily Risk Appetite Index

Variable	*Start date*	*Index loading*
Equity indices		
S&P 500 stock index	Mar. 1957	0.42
NASDAQ composite stock index	Feb. 1971	0.39
Volatility		
ICE/BofA MOVE index	Apr. 1988	–0.17
10-year Treasury note volatility (TYVIX)	May 1985	–0.15
S&P 500 volatility index (VIX)	Jan. 1990	–0.41
Bekaert-Hoerova equity variance risk premium (VRP)	Jan. 1990	–0.29
Credit spreads		
Moody's Baa corporate bond spread	Jan. 1986	–0.16
ICE/BofA US investment-grade (IG) corporate option-adjusted spread (OAS)	Jan. 1997	–0.27
ICE/BofA US high-yield (HY) corporate OAS	Jan. 1997	–0.34
3-month commercial paper (CP) spread	Apr. 1997	–0.14
J.P. Morgan emerging markets (EM) bond index (EMBI+) spread	Jan. 1998	–0.29
Bloomberg OAS for US fixed-rate mortgage-backed securities (MBS)	Aug. 2000	–0.13
Exchange rates		
US dollar exchange rate versus advanced foreign economies	Mar. 1973	–0.06
Swiss franc-Euro exchange rate	Jan. 1999	–0.17

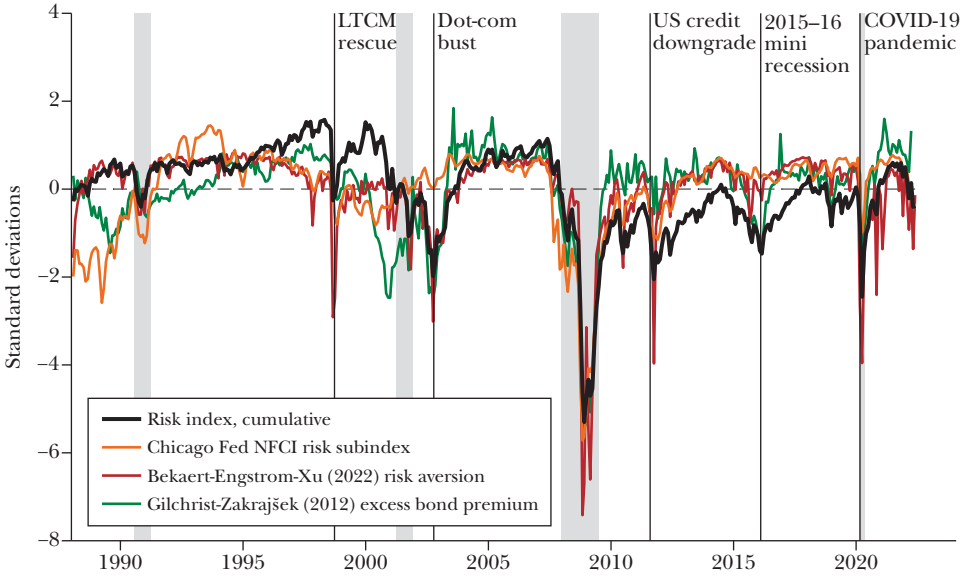
Source: For more details on sources, see the Online Appendix.
Notes: The loading column shows the weight of each variable in the index (more specifically, the components of the first eigenvector of the correlation matrix of the 14 variables). Equity indices and exchange rates are transformed as daily log returns, volatility indices are daily changes in index points, and credit spreads are daily changes in percentage points. The index is signed such that an increase corresponds to an increase in risk appetite.

Appendix describes the details of our procedure, which deals with missing observations in order to obtain a complete time series of the index. Our index accounts for about 30 percent of the common variation in the 14 component variables, which suggests a substantial amount of comovement given the variety of different assets and indicators we include.[10] We sign the index so that an increase in the index corresponds to an increase in risk appetite. The index has mean zero by construction, and we normalize it to have a standard deviation of one.

The rightmost column of Table 1 shows the loading of each variable on the index of risk appetite. Since the components are standardized, these loadings also reflect the individual contributions to the index. The sign of the loading indicates whether the variable moves in the same or the opposite direction as the index when risk appetite changes, and all loadings have the expected signs: Greater risk appetite, as measured by our index, is associated with higher equity returns, lower volatility of bond and stock returns, tighter credit spreads, and depreciation of the

[10] For comparison, the index of Miranda-Agrippino and Rey (2020) explained 21.5 percent of the variation in their panel of 858 risky asset prices.

Figure 3
Comparison of Selected Risk Indices and Market Events



Source: For details on the sources, see the Online Appendix.
Notes: All series are shown at monthly frequency, standardized to have zero mean and unit standard deviation, and signed so that an increase corresponds to an increase in risk appetite. Shading denotes recessions as dated by the National Bureau of Economic Research. Sample period is January 1988 to May 2022.

dollar and Swiss franc (the safe haven currencies). Variables related to the stock market—the two stock indexes, the VIX volatility index, and the equity variance risk premium—have the greatest influence on the index, although all the component variables have nontrivial weight.

By construction, our index captures daily changes in risk appetite. As a reality check, we can cumulate the index to produce a measure of the overall level of risk appetite at each point in time, represented by the thick black line in Figure 3. Note that because this line is the cumulation of an index with a mean of zero, it has no trend by construction. Expressing the risk appetite index in levels shows how it lines up with key historical events. As Figure 3 illustrates, large "risk-off" days—when investors' risk appetite drops—can usually be identified with specific adverse events, such as the Lehman failure in 2008, the COVID shock in 2020, and the bursting of the dotcom bubble from 2000 to 2002. The largest daily "risk-on" events are in most cases part of reversals of large "risk-off" shocks, but improvements in risk appetite can also be seen in the latter part of the 1990s, in the period between the bursting of the dot-com bubble and the beginning of the housing crisis, between the 2011 US credit downgrade and the COVID shock (with interruptions), and after the March 2020 COVID-induced financial crisis. Overall, there seems to be a pattern

of sharp drops in risk appetite followed by slow recoveries. Indeed, the largest daily changes in risk appetite are typically to the downside: Of the 25 largest changes in our sample (in absolute value), 20 were downward, and the distribution of daily changes is skewed towards large declines.[11] In contrast, on days with monetary policy FOMC announcements, changes in risk appetite were positive on average, with a mean of 0.25, and positive skewness. It appears that risk appetite behaved differently on days with FOMC announcements than on other days, with markets on average mildly reassured by the Fed, perhaps because uncertainties are resolved by the announcement (Bauer, Lakdawala, and Mueller 2022).

For comparison, Figure 3 also shows three other risk indicators from the literature, based on varying approaches, all of which span the period covered by our index: the risk aversion index developed by Bekaert, Engstrom, and Xu (2022); the risk sub-index of the Chicago Fed's NFCI; and the Gilchrist-Zakrajšek excess bond premium. These three alternative indexes give historical descriptions of risk that are qualitatively quite similar to ours. The correlations with our index are 0.60 for the Bekaert et al. index, 0.60 for the Chicago Fed risk sub-index, and 0.64 for the Gilchrist-Zakrajšek excess bond premium.[12]

Using our index, we can now examine how risk appetite is affected by monetary policy. Figure 4 shows the contemporaneous effects of a tightening surprise on the day of the FOMC announcement and the cumulative effects over the 20 trading days following the announcement, with 90-percent confidence intervals. This figure is constructed with regressions similar to those underlying Figure 2 above. On impact, the tightening surprise significantly lowers risk appetite. Over the days and weeks following the announcement, the estimated effects become larger in magnitude and even more highly statistically significant. The drift in the response of the index mirrors that of the components, some of which are also seen earlier in Figure 2. The drift in the response over the entire 20 days, measured as the difference in the 20-day response and impact response, is large and statistically significant. In fact, most of this drift occurs in the first five days after the FOMC announcement. This post-FOMC drift in risk appetite is an intriguing result that is worth further investigation.

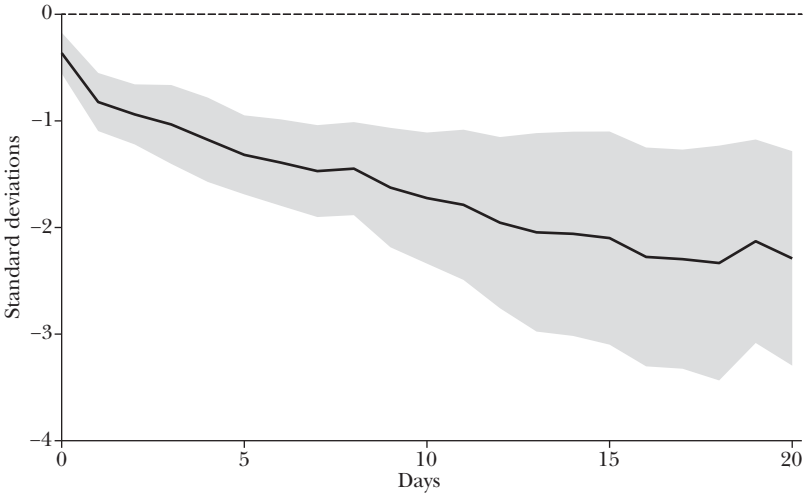
The statistical significance of the responses shown in Figure 4 is very high (for example, the t-statistic for the response after five days is 5.8). But how important economically is the response of the risk appetite index shown in Figure 4? On impact, the decline of the index to a ten basis point surprise is a little less than half its standard deviation. After ten trading days, the cumulative decline equals about 1.7 standard deviations.[13] Thus, the effects of monetary policy surprises on

[11] Our index has a skewness coefficient of –1.6, indicating the downward skew.

[12] Correlations with alternative indicators of risk, economic and financial conditions, sentiment, and uncertainty are provided in the Online Appendix, and they are also generally high.

[13] The standard deviation of eleven-day changes in the index, is higher than the standard deviation of daily changes by about the square root of eleven, because these changes are almost serially uncorrelated. The estimated decline after ten days is about half as large as this standard deviation.

Figure 4
Dynamic Response of Risk Appetite to a Surprise Monetary Tightening



Source: For details of the calculations, see the Online Appendix.
Notes: Estimated response of risk appetite index to Bauer and Swanson (2022) monetary policy surprises, scaled to a ten basis point surprise, on the day of the announcement by the FOMC (day 0) and cumulative responses over the subsequent (1–20) trading days. The risk appetite index has mean zero by construction, and we normalize it to have a standard deviation of one. The sample contains all FOMC announcements from January 1988 to December 2019. Shaded areas correspond to 90 percent confidence intervals based on Huber-White heteroskedasticity-robust standard errors.

risk appetite seem relatively large, compared to historically normal fluctuations in the index.

Figure 4 shows the response of risk appetite to unanticipated movements in current and expected values of the federal funds rate. But monetary policy actions and communication may have effects on risky asset prices and risk appetite that are not captured by event-study regressions using only monetary policy surprises. The FOMC statements and, more recently, the Fed Chair's press conferences following the release of the statement, provide additional information about various aspects of current monetary policy. These include the Fed's economic outlook, the balance of risks around the expected policy path (Bauer and Chernov forthcoming; Bauer, Lakdawala, and Mueller 2022), the policy reaction function (Bauer, Pflueger, and Sunderam 2022), the likelihood of unconventional policies (Kuttner 2018), or the likelihood of backstopping a deterioration in financial conditions (Cieslak and Vissing-Jorgensen 2021). Such information, which is not fully captured by monetary surprises based on risk-free rates with maturities of a year or less, is also likely to affect the risk appetites of investors. In the words of Kroencke, Schmeling, and Schrimpf (2021), "[M]onetary policy surprises extracted from changes in risk-free interest

rates alone will necessarily lack an important part of the information contained in monetary policy announcements."

Our estimates are consistent with the view that monetary policy actions and communication can affect risk appetite in ways not captured by monetary policy surprises: No matter how we measure these surprises or how much delay we allow for the response, we can only explain up to about 10 percent of the daily variation in risk appetite.[14] While some of the variation in risk appetite on days with FOMC announcements is certainly driven by news unrelated to monetary policy, it is hard to argue that all, or even most, of the remaining 90 percent of the daily variation in risk appetite is unrelated to monetary policy. Given the importance of monetary policy for financial markets, it seems much more plausible that these additional changes in risk appetite are due in part to the news about monetary policy on these days that is not fully reflected in the high-frequency policy surprise. This view is also supported by different types of empirical analysis, including textual analysis linking financial market reactions to the content of FOMC statements (Gardner, Scotti, and Vega 2022), stock market responses in the opposite direction from what one would expect based on monetary policy surprises (Cieslak and Schrimpf 2019; Jarociński and Karadi 2020) and, more generally, evidence on "FOMC risk shifts" (Kroencke, Schmeling, and Schrimpf 2021).

Overall, the evidence strongly suggests that monetary policy actions have pronounced effects on risk appetite in financial markets. Beyond the usual contemporaneous event-study regressions, our estimates showed that these effects build even further over the days following FOMC announcements. In the next section, we discuss our reasons for believing that the effect of the risk-taking channel on macroeconomic dynamics is substantial enough that it deserves more attention from economists and policymakers, together with the more familiar neoclassical channels for the transmission of monetary policy.

The Risk-Taking Channel and Macroeconomic Dynamics

The effects of monetary policy on asset prices and risk appetite are of independent interest, but they are only the first step in the risk-taking channel of monetary policy. The next step in the monetary transmission are the effects of changes in risk appetite and the related changes in risky asset prices on spending, employment, inflation, and other macroeconomic quantities. After all, the ultimate goal of research on the risk-taking channel of monetary policy is to understand the quantitative importance of these effects on the broader economy.

As discussed above, many existing macro-finance theories imply that changes in risk appetite are likely to play an important role in the monetary transmission via changes in asset prices, household wealth, collateral values, and intermediary

[14] For some of these alternative measures and calculations, see Table B3 in the Online Appendix.

balance sheets. In addition, empirical work on the macroeconomic effects of monetary policy suggests that changes in risk premia may be important. A prominent example is the work of Gertler and Karadi (2015), who used monetary policy surprises combined with vector autoregressions to understand the role of changes in credit costs and risk premia in monetary transmission. Their estimates show a substantial and persistent impact of monetary policy on risk premia, which is consistent with the view that changes in risk premia are an important component in the monetary transmission. Both the excess bond premium of Gilchrist and Zakrajšek (2012) and the measure of risk perceptions of Pflueger, Siriwardane, and Sunderam (2020) predict future economic activity. Other empirical studies suggest that exogenous changes in the term premium can have significant macroeconomic effects (Rudebusch, Sack, and Swanson 2007; Baumeister and Benati 2013).

Overall, both theory and evidence support the view that the risk-taking channel may be quite important for monetary transmission, and that changes in risk appetite due to monetary policy are likely to have sizeable macroeconomic effects. However, much work remains to convincingly quantify the importance of the risk-taking channel. While there is extensive evidence that monetary policy affects risk premia in financial markets, significantly less is known about how large the consequences of these effects are for economic activity and inflation. The challenge here is considerable. A full analysis would require separating the effects of monetary policy on aggregate outcomes operating through the conventional neoclassical interest rate channels mentioned at the beginning of this paper from the effects working through the risk-taking channel. Moreover, this analysis would need to take into account the arguments noted at the end of the previous section that monetary policy may affect risk appetite via channels other than the policy rate and forward guidance. One potential avenue for empirical work to address these challenges is to combine new econometric tools, including the vector autoregression methods used by Gertler and Karadi (2015) and others, supplemented with high-frequency measures of changes in risk appetite due to monetary policy announcements.[15]

Financial Stability and Optimal Monetary Policy

Some have argued that, if easy money promotes risk-taking, and if increased risk-taking in turn raises the odds of a future crisis, then monetary policy should be less aggressive in responding to downturns, effectively sacrificing near-term economic stabilization goals in the interest of longer-run financial stability (Adrian and Duarte 2016; Adrian and Liang 2018; Kashyap and Stein in this symposium). We are more agnostic on this point. While this tradeoff may be valid in principle,

[15] In preliminary work, using structural vector autoregressions with high-frequency identification, we decompose policy shocks into two components, one due to changes in risk appetite and the other due to changes in other factors. Our estimates suggest that the risk-taking channel explains a significant portion of the effects of monetary policy on output and inflation.

quantitative guidance for policymakers depends on calculation of the costs and benefits of particular strategies. We know too little about critical quantities—including the share of the variation in risk appetite attributable to monetary policy; the precise macroeconomic effects of the risk-taking channel; the relative contributions of monetary, regulatory, institutional, and other factors to bouts of financial instability; the role of initial conditions; and the long-run costs of financial instability—to do reliable cost-benefit analyses. Existing attempts to do such analyses have not provided clear answers.[16] Moreover, there may be times—perhaps following a period of crisis or recession—when the risk appetites of lenders, investors, and entrepreneurs are too low to promote healthy growth. That possibility is consistent with our evidence that most large changes in risk appetite involve greater rather than less risk aversion, so that many periods of increasing risk appetite involve a return to a normal level from below. When risk appetite is too low, more aggressive easing of monetary policy than justified by macroeconomic conditions alone could in principle be warranted. Finally, the fact that the risk-taking channel likely induces stronger effects of monetary policy on the economy than can be accounted for by neoclassical policy channels alone implies that the cost of attenuating the policy response to recessions due to financial stability concerns could be high, especially if there are alternative policy tools for dealing with financial risks.

Conclusion

The risk-taking channel of monetary policy has deservedly received increasing attention in academic and policy discussions. This article has discussed how monetary policy, via this risk-taking channel, affects both risk appetite in financial markets and macroeconomic outcomes.

There remain important questions open for future research, including the quantitative importance of the risk-taking channel for the effects of monetary policy on macroeconomic aggregates, as we have emphasized. Relatedly, a better understanding and quantification of the *mechanisms* underlying the risk-taking channel would be useful. In particular, our findings could reflect the effects of monetary surprises on the economic outlook, which in turn influence risk attitudes. Alternatively, the estimated effects could in part be the result of behavioral factors, including the reaching-for-yield phenomenon. Furthermore, our estimates only capture effects on risk appetite from unanticipated changes in the policy action, as measured by monetary policy surprises, omitting additional information provided in the statement or the chair's press conference (as well as the effects of the systematic component of monetary policy). The use of methods such as machine learning

[16] See, for example, Svensson (2017) and Guorio, Kashyap, and Sim (2018). Boyarchenko, Favara, and Schularick (2022) survey what is known about the relationship between monetary policy and financial stability, concluding that, given the variety of factors affecting stability, clear links are difficult to identify. See also Bernanke (2022, Chapter 14).

or textual analysis is one promising direction for future research about the link between policy communication and risk appetite.

The implications of the risk-taking channel for the optimal conduct of monetary policy—and in particular, for the interactions between monetary policy and financial instability—are a particularly important topic for further study. At this stage, we know too little about the effects of the risk-taking channel on both financial stability and the real economy to offer useful quantitative advice to policymakers. It is certainly possible that easier monetary policy and the resulting rise in risk appetite affects the probability and cost of a financial crisis—important unknowns in determining optimal monetary policy—but the quantitative linkages must surely depend heavily on the institutional and regulatory arrangements at a particular time and place, as well as the initial economic and financial conditions. Moreover, the behavior of our new index of risk appetite suggests that investor risk appetites are typically below normal during periods of crisis or recession. In such a situation, the tendency of monetary easing to increase risk appetites could be beneficial.

■ *We thank Aeimit Lakdawala, Mark Gertler, Maik Schmeling, Eric Swanson, Pascal Paul and Annette Vissing-Jorgensen for helpful comments, and James Lee for research assistance.*

References

Adrian, Tobias, and Fernando M. Duarte. 2016. "Financial Vulnerability and Monetary Policy." Federal Reserve Bank of New York Staff Report 804. https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr804.pdf?la=en.

Adrian, Tobias, and Nellie Liang. 2018. "Monetary Policy, Financial Conditions, and Financial Stability." *International Journal of Central Banking* 14 (1): 73–131.

Adrian, Tobias, and Hyun Song Shin. 2010. "Financial Intermediaries and Monetary Economics." In *Handbook of Monetary Economics*, Vol. 3, edited by Benjamin M. Friedman and Michael Woodford, 601–50. Amsterdam: North-Holland.

Baele, Lieven, Geert Bekaert, Koen Inghelbrecht, and Min Wei. 2020. "Flights to Safety." *Review of Financial Studies* 33 (2): 689–746.

Bansal, Ravi, and Amir Yaron. 2004. "Risks for the Long Run: A Potential Resolution of Asset Pricing Puzzles." *Journal of Finance* 59 (4): 1481–509.

Bao, Jack, Jun Pan, and Jiang Wang. 2011. "The Illiquidity of Corporate Bonds." *Journal of Finance* 66 (3): 911–46.

Bauer, Michael D., and Mikhail Chernov. Forthcoming. "Interest Rate Skewness and Biased Beliefs." *Journal of Finance*.

Bauer, Michael D., and Glenn D. Rudebusch. 2014. "The Signaling Channel for Federal Reserve Bond Purchases." *International Journal of Central Banking* 10 (3): 233–89.

Bauer, Michael D., and Eric T. Swanson. 2022. "A Reassessment of Monetary Policy Surprises and High-Frequency Identification." In *NBER Macroeconomics Annual 2022*, Vol. 37, edited by Martin Eichenbaum, Erik Hurst, and Valerie A. Ramey. Chicago: University of Chicago Press.

Bauer, Michael D., and Eric T. Swanson. Forthcoming. "An Alternative Explanation for the 'Fed Information Effect.'" *American Economic Review.*

Bauer, Michael D., Ben S. Bernanke, and Eric Milstein. 2023. "Replication data for: Risk Appetite and the Risk-Taking Channel of Monetary Policy." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E183829V1.

Bauer, Michael D., Aeimit Lakdawala, and Philippe Mueller. 2022. "Market-Based Monetary Policy Uncertainty." *Economic Journal* 132 (644): 1290–308.

Bauer, Michael D., Carolin E. Pflueger, and Adi Sunderam. 2022. "Perceptions about Monetary Policy." NBER Working Paper 30480.

Baumeister, Christiane, and Luca Benati. 2013. "Unconventional Monetary Policy and the Great Recession: Estimating the Macroeconomic Effects of a Spread Compression at the Zero Lower Bound." *International Journal of Central Banking* 9 (2): 165–212.

Becker, Bo, and Victoria Ivashina. 2015. "Reaching for Yield in the Bond Market." *Journal of Finance* 70 (5): 1863–902.

Bekaert, Geert, Eric C. Engstrom, and Steven R. Grenadier. 2010. "Stock and Bond Returns with Moody Investors." *Journal of Empirical Finance* 17 (5): 867–94.

Bekaert, Geert, Eric C. Engstrom, and Nancy R. Xu. 2022. "The Time Variation in Risk Appetite and Uncertainty." *Management Science* 68 (6): 3975–4004.

Bekaert, Geert, and Marie Hoerova. 2014. "The VIX, the Variance Premium and Stock Market Volatility." *Journal of Econometrics* 183 (2): 181–92.

Bekaert, Geert, Marie Hoerova, and Marco Lo Duca. 2013. "Risk, Uncertainty and Monetary Policy." *Journal of Monetary Economics* 60 (7): 771–88.

Bernanke, Ben S. 2007. "The Financial Accelerator and the Credit Channel." Speech, Federal Reserve Bank of Atlanta Credit Channel of Monetary Policy in the Twenty-First Century Conference, Atlanta, June 15, 2007. https://www.federalreserve.gov/newsevents/speech/bernanke20070615a.htm.

Bernanke, Ben S. 2022. *21st Century Monetary Policy: The Federal Reserve from the Great Inflation to COVID-19.* New York: W. W. Norton & Company.

Bernanke, Ben S., and Mark Gertler. 1995. "Inside the Black Box: The Credit Channel of Monetary Policy Transmission." *Journal of Economic Perspectives* 9 (4): 27–48.

Bernanke, Ben S., Mark Gertler, and Simon Gilchrist. 1999. "The Financial Accelerator in a Quantitative Business Cycle Framework." In *Handbook of Macroeconomics*, Vol. 1C, edited by John B. Taylor and Michael Woodford, 1341–93. Amsterdam: North-Holland.

Bernanke, Ben S., and Kenneth N. Kuttner. 2005. "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *Journal of Finance* 60 (3): 1221–57.

Boivin, Jean, Michael T. Kiley, and Frederic S. Mishkin. 2010. "How Has the Monetary Transmission Mechanism Evolved over Time?" In *Handbook of Monetary Economics*, Vol. 3, edited by Benjamin M. Friedman and Michael Woodford, 369–422. Amsterdam: North-Holland.

Borio, Claudio, and Haibin Zhu. 2012. "Capital Regulation, Risk-Taking and Monetary Policy: A Missing Link in the Transmission Mechanism?" *Journal of Financial Stability* 8 (4): 236–51.

Boyarchenko, Nina, Giovanni Favara, and Moritz Schularick. 2022. "Financial Stability Considerations for Monetary Policy: Empirical Evidence and Challenges." Finance and Economics Discussion Series 2022-006, Federal Reserve Board.

Brave, Scott, and R. Andrew Butters. 2011. "Monitoring Financial Stability: A Financial Conditions Index Approach." *Federal Reserve Bank of Chicago Economic Perspectives* 35 (1-2): 22–43.

Bruno, Valentina, and Hyun Song Shin. 2015. "Capital Flows and the Risk-Taking Channel of Monetary Policy." *Journal of Monetary Economics* 71: 119–32.

Campbell, John Y., and Roman Sigalov. 2022. "Portfolio Choice with Sustainable Spending: A Model of Reaching for Yield." *Journal of Financial Economics* 143 (1): 188–206.

Campbell, John Y., and John H. Cochrane. 1999. "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior." *Journal of Political Economy* 107 (2): 205–51.

Chari, Anusha, Karlye Dilts Stedman, and Christian Lundblad. 2020. "Capital Flows in Risky Times: Risk-on/Risk-off and Emerging Market Tail Risk." NBER Working Paper 27927.

Chan, Yeung Lewis, and Leonid Kogan. 2002. "Catching Up with the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices." *Journal of Political Economy* 110 (6): 1255–85.

Cieslak, Anna, and Andreas Schrimpf. 2019. "Non-monetary News in Central Bank Communication."

Journal of International Economics 118: 293–315.

Cieslak, Anna, and Annette Vissing-Jorgensen. 2021. "The Economics of the Fed Put." *Review of Financial Studies* 34 (9): 4045–89.

Cochrane, John H. 2005. *Asset Pricing*. Revised ed. Princeton, NJ: Princeton University Press.

Coudert, Virginie, and Mathieu Gex. 2008. "Does Risk Aversion Drive Financial Crises? Testing the Predictive Power of Empirical Indicators." *Journal of Empirical Finance* 15 (2): 167–84.

Datta, Deepa, Juan M. Londono, Bo Sun, Daniel O. Beltran, Thiago Ferreira, Matteo Iacoviello, Mohammad R. Jahan-Parvar, Canlin Li, Marius del Giudice Rodriguez, and John Rogers. 2017. "Taxonomy of Global Risk, Uncertainty, and Volatility Measures." International Finance Discussion Paper 1216.

Disyatat, Piti. 2011. "The Bank Lending Channel Revisited." *Journal of Money, Credit and Banking* 43 (4): 711–34.

Drechsler, Itamar, Alexi Savov, and Philipp Schnabl. 2018. "Liquidity, Risk Premia, and the Financial Transmission of Monetary Policy." *Annual Review of Financial Economics* 10: 309–28.

Forbes, Kristin J., and Francis E. Warnock. 2021. "Capital Flow Waves—or Ripples? Extreme Capital Flow Movements since the Crisis." *Journal of International Money and Finance* 116: 102394.

Gai, Prasanna, and Nicholas Vause. 2006. "Measuring Investors' Risk Appetite." *International Journal of Central Banking* 2(1): 167-88.

Gardner, Ben, Chiara Scotti, and Clara Vega. 2022. "Words Speak as Loudly as Actions: Central Bank Communication and the Response of Equity Prices to Macroeconomic Announcements." *Journal of Econometrics* 231 (2): 387–409.

Gârleanu, Nicolae, and Lasse Heje Pedersen. 2011. "Margin-Based Asset Pricing and Deviations from the Law of One Price." *Review of Financial Studies* 24 (6): 1980–2022.

Gertler, Mark, and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7 (1): 44–76.

Gilchrist, Simon, David López-Salido, and Egon Zakrajšek. 2015. "Monetary Policy and Real Borrowing Costs at the Zero Lower Bound." *American Economic Journal: Macroeconomics* 7 (1): 77–109.

Gilchrist, Simon, and Egon Zakrajšek. 2012. "Credit Spreads and Business Cycle Fluctuations." *American Economic Review* 102 (4): 1692–720.

Gourio, François, Anil K. Kashyap, and Jae W. Sim. 2018. "The Trade Offs in Leaning Against the Wind." *IMF Economic Review* 66 (1): 70–115.

Gürkaynak, Refet S., Brian Sack, and Eric T. Swanson. 2005. "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking* 1 (1): 55–93.

Hanson, Samuel G., and Jeremy C. Stein. 2015. "Monetary Policy and Long-Term Real Rates." *Journal of Financial Economics* 115 (3): 429–48.

Hanson, Samuel G., David O. Lucca, and Jonathan H. Wright. 2021. "Rate-Amplifying Demand and the Excess Sensitivity of Long-Term Rates." *Quarterly Journal of Economics* 136 (3): 1719–81.

He, Zhiguo, and Arvind Krishnamurthy. 2013. "Intermediary Asset Pricing." *American Economic Review* 103 (2): 732–70.

Jarocinski, Marek, and Peter Karadi. 2020. "Deconstructing Monetary Policy Surprises—The Role of Information Shocks." *American Economic Journal: Macroeconomics* 12 (2): 1–43.

Kekre, Rohan, and Moritz Lenel. 2022. "Monetary Policy, Redistribution, and Risk Premia." *Econometrica* 90 (5): 2249–82.

Kroencke, Tim A., Maik Schmeling, and Andreas Schrimpf. 2021. "The FOMC Risk Shift." *Journal of Monetary Economics* 120: 21–39.

Kuttner, Kenneth N. 2001. "Monetary Policy Surprises and Interest Rates: Evidence from the Fed Funds Futures Market." *Journal of Monetary Economics* 47 (3): 523–44.

Kuttner, Kenneth N. 2018. "Outside the Box: Unconventional Monetary Policy in the Great Recession and Beyond." *Journal of Economic Perspectives* 32 (4): 121–46.

Lian, Chen, Yueran Ma, and Carmen Wang. 2019. "Low Interest Rates and Risk-Taking: Evidence from Individual Investment Decisions." *Review of Financial Studies* 32 (6): 2107–48.

Miranda-Agrippino, Silvia, and Hélène Rey. 2020. "U.S. Monetary Policy and the Global Financial Cycle." *Review of Economic Studies* 87 (6): 2754–76.

Miranda-Agrippino, Silvia, and Hélène Rey. 2022. "The Global Financial Cycle." In *Handbook of International Economics*, Vol. 6, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 1–43. Amsterdam: North-Holland.

Nakamura, Emi, and Jón Steinsson. 2018a. "Identification in Macroeconomics." *Journal of Economic Perspectives* 32 (3): 59–86.

Nakamura, Emi, and Jón Steinsson. 2018b. "High-Frequency Identification of Monetary Non-neutrality: The Information Effect." *Quarterly Journal of Economics* 133 (3): 1283–330.

Paligorova, Teodora, and João A. C. Santos. 2017. "Monetary Policy and Bank Risk-Taking: Evidence from the Corporate Loan Market." *Journal of Financial Intermediation* 30: 35–49.

Pflueger, Carolin, Emil Siriwardane, and Adi Sunderam. 2020. "Financial Market Risk Perceptions and the Macroeconomy." *Quarterly Journal of Economics* 135 (3): 1443–91.

Rudebusch, Glenn D., Brian P. Sack, and Eric T. Swanson. 2007. "Macroeconomic Implications of Changes in the Term Premium." *Federal Reserve Bank of St. Louis Review* 89 (4): 241–69.

Ramey, Valerie A. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics*, Vol. 2, edited by John B. Taylor and Harald Uhlig, 71–162. Amsterdam: North-Holland.

Svensson, Lars E. O. 2017. "Cost-Benefit Analysis of Leaning against the Wind." *Journal of Monetary Economics* 90: 193–213.

Wachter, Jessica A. 2013. "Can Time-Varying Risk of Rare Disasters Explain Aggregate Stock Market Volatility?" *Journal of Finance* 68 (3): 987–1035.

Journal of Economic Perspectives—Volume 37, Number 1—Winter 2023—Pages 101–120

Landings, Soft and Hard: The Federal Reserve, 1965–2022

Alan S. Blinder

When the Federal Reserve set out to reduce inflation in March 2022, talk of "hard" versus "soft" landings was in the air. What does it all mean? And what does US monetary history teach us about the likelihood of achieving a soft landing?

Start with the basics. When the Fed gears up to fight either actual or incipient inflation, it generally tightens monetary policy by raising the interest rate in the federal funds market, the overnight market for uncollateralized lending of bank reserves. This step then increases other interest rates throughout the economy via financial arbitrage—and also reduces stock prices and, most likely, raises the dollar exchange rate. Higher interest rates, of course, make credit more expensive and dampen aggregate demand, especially for houses and motor vehicles. The central idea is to slow down the growth of aggregate demand, thereby taking some of the inflationary "steam" out of the economy.

If the Federal Open Market Committee manages to take out just enough steam, but not too much, it might engineer a "soft landing," with inflation stabilized or reduced and either no recession or a negligible one. More monetary tightening than that, however, will precipitate a recession—a "hard landing." And too little tightening would fail to cure the inflation problem. The Fed is thus operating within a narrow corridor and should probably be thinking constantly about Goldilocks and the three bears.

■ *Alan S. Blinder is the Gordon S. Rentschler Memorial Professor of Economics and Public Affairs, Princeton University, Princeton, New Jersey.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.1.101.

Soft landings are the holy grail of monetary policy. Conventional wisdom holds that they are nearly as elusive—something that central banks rarely achieve. Why are they so difficult? Perhaps the main reason is what Milton Friedman (1961) famously called the "long and variable lags" between monetary policy decisions and their effects on the economy. The Federal Open Market Committee may feel it has reasonable estimates of the monetary policy multiplier—how much GDP growth will be slowed by, say, raising the federal funds rate 100 basis points. But the timing of the response is difficult, maybe impossible, to predict. When the Fed raises the funds rate, other nominal interest rates across the economy tend to rise also, but not immediately nor in lock-step. Next, firms and household must adjust their borrowing and spending plans in response to these higher nominal interest rates. Such reactions will be influenced strongly by expectations of how the economy will evolve, including expectations of future rates of inflation and future real interest rates. Finally, those decisions by households and firms will have multiplier effects as they percolate through the rest of the economy. The overall process from altering the federal funds rate until its full effects on, say, GDP can easily take a year or two. As Friedman said, *long*—and also difficult to predict.

These long and variable lags create a very real human danger: that monetary policymakers, seeing little or no immediate effect in reducing inflation, will keep raising interest rates beyond what is actually needed to reduce inflation, eventually causing a hard landing. (Conversely, a central bank seeking to stimulate the economy may keep loosening monetary policy for too long, eventually leading to inflation. But this paper focuses on tightenings.) This danger, I believe, is the major reason why Rudiger Dornbusch (1997) famously opined years ago, "No postwar recovery has died in bed of old age—the Federal Reserve has murdered every one of them."

Another major hazard to achieving a soft landing comes from the external environment, especially from adverse supply shocks. Even if a central bank seeking to reduce inflation does everything right, external shocks—like the oil- and food-price hikes precipitated by the Russian invasion of Ukraine in 2022—can intervene and upset the best-laid plans. Thus, to achieve a soft landing, a central bank, having little or no control over aggregate supply, must be *both* lucky *and* good.

The conventional wisdom seems to be that the Federal Reserve has managed to pull off only one soft landing in its history: the tightening cycle that began in February 1994 and ended in February 1995. But is that assessment correct? This paper examines the historical record and concludes that it is not. Prior to the tightening that began in 2022, the Federal Reserve has tightened monetary policy to combat inflation eleven previous times since 1965.[1] I will argue that soft and "softish" landings have been more common than is generally believed. But first, some conceptual preliminaries.

[1] The choice of starting year, 1965, and some of the historical analysis derive from my new book, Blinder (2022).

Goals and Definitions

Any serious historical investigation of when and why tighter monetary policies have led to either soft or hard landings must begin with an understanding of the Federal Reserve's goals, as well as operational definitions of both "tightening" and "soft landing."

The Dual Mandate of the Federal Reserve

The Federal Reserve's monetary policy has long been guided by a dual mandate to pursue both "maximum employment" and "stable prices"—two vague phrases that appear in the Federal Reserve Act, as amended. Formal models of how the Fed might balance these two goals, and steer a course between them, share the basic idea that the central bank has a target level of inflation in mind.[2] When inflation exceeds—or, perhaps, *is expected to* exceed—its target, the central bank will, *ceteris paribus*, raise its policy interest rate to reduce aggregate demand. However, the dual mandate implies an inherent interest in keeping the "landing" as "soft" as possible—that is, reducing inflation without doing too much damage to real output and employment. Conversely, when the Fed falls short of its "maximum employment" target, it will reduce interest rates, seeking to do so without causing higher inflation.

Since 2012, but not before, the Federal Reserve (2012) has set an official target of 2 percent inflation as measured by the price deflator for personal consumption expenditures (PCE). Operationally, the Fed focuses more on the "core" PCE deflator, that is, the inflation rate after stripping out the effects of food and energy prices—over which monetary policy has little or no influence. Since 2020, but not before, the Federal Reserve (2020) has stated that this target is to be achieved *on average* over an unspecified period of time—so-called "flexible average inflation targeting." For several years prior to its landmark 2012 decision to officially embrace a numerical inflation target for the first time, the Fed was widely believed to be targeting core inflation as measured by the Consumer Price Index in the 1.5–2 percent range (Shapiro and Wilson 2022).[3] The Federal Open Market Committee never officially adopted that target, but neither did it try to disabuse markets of that belief.

[2]For example, a model of a central bank's loss function might be:

$$L_t = (\pi_t - \pi^*)^2 + \lambda(y_t - y_t^*)^2 \quad (\lambda > 0)$$

where π is the inflation rate, π^* is the central bank's target inflation rate, y is output, and y^* is potential output. Alternatively, a "Taylor rule" formulation of monetary policy can be written as

$$i_t = r^* + \pi_t + \rho(\pi_t - \pi^*) + \beta(y_t - y_t^*) \quad (\rho, \beta > 0)$$

where i is the central bank's (*nominal*) policy rate and r^* is the neutral *real* interest rate. Perhaps curiously, equations like these are also typically employed to analyze the monetary policies of central banks that do not have dual mandates, that is, which claim to be inflation targeters. But my focus here is on the Federal Reserve.

[3]The deflators as measured by the Consumer Price Index and the Personal Consumption Expenditures index differ in myriad ways (for discussion, see Johnson 2017). Historically, CPI inflation averages about 0.5 percent higher than PCE inflation over long periods of time. I focus more on CPI inflation in this

Going back further, into the 1990s and early 2000s, the Federal Open Market Committee probably did have a *very* unofficial, indeed secretive, 2 percent target for inflation as measured by the Consumer Price Index. At an important July 1996 meeting of the Federal Open Market Committee, now famous among connoisseurs of this subject, then-Governor Janet Yellen (the dove) and then-President of the Richmond Fed Alfred Broaddus (the hawk) debated what the Fed's long-term inflation target should be. There seemed to be a consensus that day—though certainly unofficial—that 2 percent inflation in the Consumer Price Index was about right. But at the end of the debate, Fed Chair Alan Greenspan warned the committee that this "consensus" must be kept secret: "If the 2 percent inflation figure gets out of this room, it is going to create more problems for us than I think any of you might anticipate."[4]

Prior to that 1996 meeting, there was probably no clear inflation target even *inside* the Federal Open Market Committee, much less *outside* it. For example, I can testify from personal experience as the Fed's vice-chair at the time that there was no agreement on the inflation target during the highly successful 1994–1995 tightening cycle. Rather, opinions about the target seemed to vary across committee members between zero and 3 percent (the latter was the actual inflation rate at the time). Nonetheless, over the decades, Fed policymakers have reacted to excessively high inflation by raising interest rates—with or without a specific inflation target. They knew high inflation when they saw it.

What is Monetary Tightening?

Defining a monetary policy tightening is *almost* straightforward. Although the Fed has employed various instruments and intermediate targets throughout its history, its operations most of the time have focused on the federal funds rate (Bernanke and Blinder 1992)—the interest rate prevailing in the overnight lending market for uncollateralized bank reserves, where the lenders and borrowers are primarily depository institutions.

Throughout this paper, I use the *effective federal funds rate*, as archived in the FRED database maintained by the Federal Reserve Bank of St. Louis. When this interest rate rises by a sizable amount over a protracted period, that is a tightening cycle. Figure 1 displays the history of the effective federal funds rate since 1965, with numbers from 1 to 11 indicating the peaks of the eleven tightening cycles studied here.

Three points about, and reasons for choosing, the effective federal funds rate are worth making here. First, while the Federal Open Market Committee announced a crisp numerical target, precise to the basis point, for the federal funds

paper because it is the measure that gets public attention. Furthermore, the Fed's focus on PCE inflation dates only from 2012.

[4]See the transcript for the July 1996 meeting at https://www.federalreserve.gov/monetarypolicy/files/FOMC19960703meeting.pdf. Transcripts of meetings of the Federal Open Market Committee are released with a five-year lag. They are available at https://www.federalreserve.gov/monetarypolicy/fomc_historical.htm.

Figure 1
The Effective Federal Funds Rate, 1965–2021

Source: Board of Governors of the Federal Reserve System
Note: The effective federal funds rate is a volume-weighted median of actual overnight federal funds transactions. Shaded areas show recessions. The numbers show peaks of 11 monetary tightening cycles. Data are monthly.

rate (almost) consistently from February 1994 until December 2008, it did not do so before or after. Since December 2008, the Committee has generally announced a 25 basis-point *range* as its target. In the years prior to 1994, its target range for the federal funds rate was often not announced at all, and statements made about the target were sometimes as wide as 400 or 500 basis points. Here is one example from the policy directive of the January 30–31, 1984 meeting:[5]

> The Chairman may call for Committee consultation if it appears . . . that pursuit of the monetary objectives and related reserve paths during the period before the next meeting is likely to be associated with a federal funds rate persistently outside a range of 6 to 10 percent.

With a range that wide, the federal funds rate target is not well defined.

Second, even using the *effective* rate requires some judgment whenever the effective rate deviates slightly from the Fed's stated *target* rate, which occurs frequently.

[5]This quotation from the Federal Open Market Committee and others like it come from the Fed's website, http://federalreserve.gov, either from FOMC statements or from verbatim transcripts.

The beginning of the celebrated soft landing of 1994–1995 provides a fine example. On February 4, 1994, Chair Alan Greenspan announced "that the Federal Open Market Committee decided to increase slightly the degree of pressure on reserve positions. The action is expected to be associated with a small increase in short-term money market interest rates." Cryptic, for sure. But Fed watchers at the time understood that Greenspan meant 25 basis points—from a 3 percent federal funds rate target to 3.25 percent. If you look at the monthly average effective federal funds rate, however, you find that it rose from 2.96 percent in December 1993 to 3.05 percent in January 1994 and then to 3.25 percent in February 1994. For consistency with the episodes before and after 1994, I date the beginning of this tightening cycle as December 1993 and take the starting federal funds rate to be 2.96 percent. Similar deviations occur around other peaks and troughs. While other observers might make slightly different judgments on timing, nothing much hinges on such decisions.

Third, the federal funds rate has at times bounced up or down for a while but then reversed course, making it difficult to select specific monthly peaks and troughs. For example, consider the tumultuous late 1970s and early 1980s. Numbers 4 and 5 in the graph indicate where I have dated interest rate peaks: April 1980 and January 1981. While the former is a clear peak, the latter could just as well have been dated June 1981, because the federal funds rates in January and June were almost identical. Similarly, I do not label April 1982 as another rate peak, despite the 250 basis point rise in the federal funds rate between December 1981 and April 1982, because the Fed was grappling (unsuccessfully) with the application of monetarist doctrine during this period while the economy was cratering. It therefore seems more natural to view the entire period from June 1981 into the early months of 1983 as one long easing cycle—and to interpret the four months from December 1981 to April 1982 as an aberration probably induced by confusion over whether, when, and how to follow or step away from the monetarist policy of constant money growth.

Landings: Soft, Softish, and Hard

The final terminological issue is how to distinguish between a soft and hard landing. One option is to use a binary definition of hard versus soft, depending on whether or not a recession ensued according to the official dates selected by the Business Cycle Dating Committee of the National Bureau of Economic Research (available at https://www.nber.org/research/business-cycle-dating). However, I prefer to think of a continuum ranging from the perfect soft landing of 1995 to the extremely hard landing of 1981–1982.

In the narrative that follows, I use real GDP, the civilian unemployment rate, and the NBER business cycle dates as indicators of real activity. If GDP declines by less than 1 percent or there is no NBER recession for at least a year after a Fed tightening cycle, I will call that a "softish" landing. As we will see, several Fed landings qualify as "softish" by this definition.

I now take up the eleven monetary tightening episodes in chronological order, following the numbering in Figure 1. For impatient readers, a far-too-quick summary is offered in Table 1, near the end of the paper.

Episode 1: September 1965–November 1966

Understanding this first episode requires a detour into fiscal policy. By mid-1965, with the economy near full employment (an unemployment rate of about 4.5 percent), it was apparent that spending on the Vietnam War was about to surge. Even though core inflation as measured by the Consumer Price Index was a mere 1.5 percent at the time, the chair of the Fed, William McChesney Martin, and his colleagues on the Federal Open Market Committee were worried about incipient inflationary pressures from an overheated economy—as were members of the President's Council of Economic Advisers, as evidenced by the eyewitness accounts of Walter Heller (1966, p. 85ff) and Arthur Okun (1970, p. 71). These worries proved well-founded; real GDP growth, powered by defense spending, averaged a sizzling 8.8 percent over the five quarters 1965:I through 1966:I, inclusive. The core inflation rate for the Consumer Price Index rose steadily from January through November 1966.

Labeling this first episode as a tightening may nonetheless be controversial. After all, the federal funds rate rose by only about 175 basis points during this period and then quickly declined by about as much. Reis (2021) has recently criticized Fed Chair Martin for losing the inflation anchor by giving up on monetary tightening too soon. But Okun (1970, p. 81), who chaired the Council of Economic Advisers at the time, rendered a much kinder contemporaneous judgment: "The Federal Reserve Board put on an outstanding performance in 1966, making wise judgments and, most of all, having the courage to act promptly and decisively on them." I am more inclined toward Okun's view. Besides, the rate-raising cycle of 1965–1966 lasted over a year, which is a fairly typical length for a tightening, precipitated a severe cutback in credit known as the "credit crunch" (discussed in Burger 1969), and brought on what was then called a "growth recession"—that is, a sharp slowdown from rapid GDP growth to below-trend GDP growth.

President Lyndon B. Johnson and Fed Chair Martin certainly viewed the Fed's policy response in 1965–1966 as a serious tightening of monetary policy. Indeed, it precipitated a sharp confrontation between the two men, during which Johnson inquired whether he had the authority to fire Martin (he did not). The president was especially infuriated when Martin's Fed "rang the gong" by raising the discount rate by 50 basis points on December 6, 1965. In those days, the federal funds rate was neither announced nor salient; it was discount rate announcements that grabbed everyone's attention—including the president's (for more detail on this episode, see Blinder 2022, pp. 18–20).

Johnson need not have worried, however, as the landing was soft. GDP growth decelerated sharply to just 2.7 percent over the last three quarters of 1966, and the core Consumer Price Index inflation rate stabilized in the 3–3.5 percent range. The unemployment rate, which stood at 4.3 percent at the beginning of this tightening cycle, was actually down to 3.6 percent at the end. There was no recession. Soft for sure. But was it a landing, or only the first act of a longer episode? My judgment is that it should count as a soft landing, but there is room for disagreement.

Episode 2: July 1967–August 1969

The main reason for disagreement is that, after stabilizing for about a year, inflation began to rise again in December 1967, reaching about 6 percent by the spring of 1969, an astounding figure at the time. The Fed saw higher inflation coming and began what turned out to be a two-year tightening cycle in July 1967—notice the early start date. The federal funds rate would eventually rise by about 540 basis points. But that was only enough to stabilize inflation around 6 percent, not to pull it down.

For what was probably the first and last time in US history, fiscal policy pitched in deliberately to try to slow down the growth of aggregate demand. (There were subsequent fiscal contractions, but they were motivated by shrinking the budget deficit, not by slowing growth.) It took a lot of persuasion from his economic advisers, but President Johnson was finally convinced to ask Congress for a tax hike in January 1967. Again, notice the early date—six months *before* the Fed started tightening, which in turn was five months before core inflation started rising. It is remarkable to the modern eye, but in those days fiscal policy was often thought of as the first line of defense against inflation. Consider these stunning words from the February 1968 *Economic Report of the President* (CEA 1968, pp. 84–85): "[I]t has been and remains the conviction of both the Administration and the Federal Reserve System that the Nation should depend on fiscal policy, not monetary policy, to carry the main burden of the additional restraint on the growth of demand that now appears necessary for 1968."

Yes, you read that right: "*fiscal* policy, not *monetary* policy." Thus, while the Fed did raise the federal funds rate by 540 basis points, it was playing second fiddle. In first fiddle, it took almost 18 months for President Johnson to persuade Congress to pass the Revenue and Expenditure Control Act of 1968, a fiscal contraction of about 2 percent of GDP whose centerpiece was a 10 percent *temporary* (and hence less powerful) income tax surcharge. By that time, aggregate demand had gathered a full head of inflationary steam.

When inflation did not decline promptly, the 1968 tax surcharge was quickly branded a failure. Monetary policy shared the blame. But the growth rate of real GDP growth did decline, from a 5.6 percent annual rate in the four quarters ending 1968:II to a 3 percent annual rate over the following five quarters. An NBER recession began in January 1970,[6] and growth was negative in three of the next five quarters. But the *net* cumulative decline in real GDP was a mere 0.6 percent, which seems pretty "soft." The unemployment rate, however, rose from 3.8 percent when the Fed tightening started, to 6.1 percent in August 1971. Not so soft.

In sum, fiscal *and* monetary policy tightenings *together* in 1967–1969 brought on a recession in 1970. But inflation did not begin to decline for another year, which is squarely in line with modern estimates of the lag between monetary policy and

[6]The NBER peak month is December 1969. Here and throughout the paper, I normally refer instead to the first month of *recession*—in this case, January 1970.

inflation. Those estimates were not available in the late 1960s, however. So *contemporaneously*, the whole effort was viewed as a policy failure. For purposes of this paper, I count the episode as a "softish" landing which brought core inflation down to about 4.5 percent by the time President Richard Nixon invoked wage-and-price controls in 1971.

Episode 3: February 1972–July 1974

The years 1972–1974 were tumultuous for the US economy, featuring three big events with major effects on inflation. First, President Nixon invoked wage-and-price controls in August 1971, which were dismantled in stages mainly in 1973 and 1974—then causing a surge in inflation. Second, there was huge monetary policy stimulus leading up to the 1972 presidential election under Fed Chair Arthur Burns, with strong encouragement from Nixon (Abrams 2006). Third, the first waves of severe supply shocks occurred in 1972–1974. As inflationary factors, the latter two were probably far more powerful than the price controls, and together they raised the core inflation rate from 3.3 percent in February 1972 to 8.8 percent by July 1974; headline inflation rose even more, from 3.8 to 11.5 percent over that same period. The last of these (the supply shocks) was a clear case of bad luck.

The waxing and waning of these three inflation shocks probably had far more influence over price behavior during this period than anything the Federal Reserve did. That said, the Federal Open Market Committee was far from idle. It raised the federal funds rate by about 960 basis points between February 1972 and July 1974, and—probably by no coincidence—the economy cratered. The recession of 1973–1975 was the longest and deepest in post–World War II history up to that point. It was certainly a hard landing.

To understand what the Fed did, it is useful to note that the Fed's previous tightening cycle, which ended in August 1969, had left the federal funds rate at 9.2 percent, or almost 4 percentage points above inflation. That seemingly high real rate was not far from the estimate that Holston, Laubach, and Williams (2017) would later make of the neutral real interest rate at that time, although there was little to no contemporaneous talk about the *real* federal funds rate in 1969. The ensuing period through February 1972, however, assisted somewhat by the negative impact of price controls on inflation,[7] had allowed the federal funds rate to drift all the way down to 3.3 percent—a bit below the inflation rate. Doing its part to fight the war on inflation, the Federal Open Market Committee started raising rates in February 1972. However, the real federal funds rate remained in negative territory, making monetary policy decidedly stimulative. A year later, however, the federal funds rate was up to 6.6 percent, headline inflation was still

[7] Blinder and Newton (1981) estimated that price controls reduced the annual Consumer Price Index inflation rate by about 1.2–1.5 percentage points between July 1971 and February 1974, and that most of that reduction came rushing back between February and October 1974.

3.9 percent, and core inflation had fallen to 2.8 percent. By that measure, money was tight again, although the 12-month growth rate of the M2 money supply—which got more attention than the federal funds rate at the time—was still about 12 percent.

Then, both interest rates and inflation—fed at first mostly by food shocks—began to explode. Between February and August 1973, the headline inflation rate leaped from 3.9 percent to 7.4 percent, leading President Nixon to impose a second (and less effective) freeze on price increases. The Federal Open Market Committee responded by raising the federal funds rate from 6.6 percent all the way up to 10.4 percent over these six months. Significantly, but mostly unnoticed at the time, the core inflation rate barely budged over the period, rising only from 2.8 percent in February to 3.2 percent in August. The oil price shocks had yet to occur, but we were already firmly in the grip of supply shocks, with core and headline measures diverging sharply.

In the fall of 1973, stagflation really took off when the Organization of the Petroleum Exporting Countries (OPEC) nearly quadrupled the price of oil within just three months. Prices of energy-related goods and services naturally followed suit, and headline inflation skyrocketed from 7.4 percent in September 1973 to 10 percent in February 1974 and then to a peak of 12.2 percent in November 1974. Core inflation rose even faster, helped along by the end of price controls. It soared from just 3.8 percent in September 1973 to 11.2 percent in November 1974.

It is hard for modern readers to get inside the heads of Fed Chair Arthur Burns and other central bankers at the time. Stagflation was a new and puzzling phenomenon. History had taught them that inflation and unemployment moved in opposite directions. Recessions were supposed to curb inflation. Yet there they were, watching an inflationary recession unfold. The now-familiar dilemma posed by adverse supply shocks was novel at the time. How should they react? Tighten monetary policy to combat inflation, which was the usual response? Or ease policy to limit the recession—also the usual response?

It was perhaps not surprising that the Fed and other central banks vacillated. As we have just seen, the federal funds rate rose sharply to fight inflation between February 1972 and September 1973, reaching a peak of 10.8 percent. But then, as the economy weakened in the period 1973:3–1974:1, the Fed backed off, lowering the federal funds rate to 9 percent by February 1974. After that, the rate soared again, to 12.9 percent in July 1974, as inflation rose into double digits. But then it went down to 5.2 percent by May 1975. At that point, the unemployment rate had reached a frightening 9 percent—the highest rate then recorded since the 1930s. Yet inflation remained elevated at 9.3 percent for the headline Consumer Price Index and 10.5 percent for core inflation.

The long, deep recession eventually did its work, and those two inflation rates were down to 7.1 percent and 6.7 percent by December 1975. The landing was extremely hard, however. The Fed encountered a serious bout of bad luck in this episode, but also had no clear vision of how to cope with supply shocks.

Episode 4: January 1977–April 1980

By the end of 1976, headline Consumer Price Index inflation drifted down to as low as 5 percent, with core inflation at about 6 percent. While those rates were high by US historical standards, they looked pretty good compared to the double-digit rates of just a year or two earlier. That uneasy peace with inflation lasted only another year, however, and then started to unravel in 1978, largely due to another round of shocks to global food prices.

In the United States, the growth rate of the food and beverages component of the Consumer Price Index, which had dropped to a low of 1 percent (seasonally adjusted annual rate) in December 1976, began rising, eventually peaking at about 13 percent in February 1979. With a relative importance in the Consumer Price Index of about 18 percent, those rising food prices alone contributed about 2 percentage points to headline inflation in 1978 and 1.75 points in both 1979 and 1980 (Blinder and Rudd 2013, p. 141).[8] While President Jimmy Carter and his Fed Chair, G. William Miller, absorbed much of the blame, Rudd and I, in reviewing this bit of history, remarked that "we are deeply skeptical that agricultural diseases, bad weather, and the hog cycle were lagged effects of monetary policy" (Blinder and Rudd 2013, p. 142).

That said, it was oil prices, not food prices, that dominated the headlines in 1979 and 1980. What is now called "OPEC II" hit the world economy when the Iranian Revolution of 1978–1979, followed by Iraq's invasion of Iran, sent crude oil prices skyrocketing. The pass-through into consumer prices was quick and dramatic. The energy component of the Consumer Price Index, which registered just 7 percent inflation (seasonally adjusted annual rate) in November 1978, soared to 47 percent by March 1980. By then, the 12-month headline inflation rate was 14.6 percent, while core inflation was "just" 12.5 percent.

Myth has it that the Fed was asleep at the wheel in the waning days of Fed Chair Arthur Burns, who once infamously declared inflation to be beyond the Fed's control (Burns 1979), and throughout the brief chairmanship of Miller (March 1978–August 1979), who was clearly ill-suited to the job—by his own admission (Meltzer 2009, p. 923). But this particular myth is not true. The federal funds rate rose from a January 1977 low of 4.6 percent up to 6.8 percent in Burns's final month (February 1978). During the Miller interlude, it rose all the way to 10.9 percent before the indomitable Paul Volcker took over. Those interest rates were not

[8]In addition, a severe measurement error in the cost of homeownership exaggerated inflation in the years 1978–1981. Before 1983, the US Bureau of Labor Statistics treated the mortgage interest rate as a *price*. Thus, when interest rates rose, whether because of inflation or the Fed, the Consumer Price Index inflation rate mechanically rose, too. When interest rates were relatively low in the 1960s and early 1970s, this practice did not make a large difference to overall inflation rates. However, during the four years 1978–1981, this flaw in the index averaged about 1.3 percentage points. (I calculated that number from the underlying data in the CPI Detailed Reports. These data, being ancient by now, are hard to find. I am grateful to Judd Cramer for pointing me to the St. Louis Fed's FRASER database: https://fraser.stlouisfed.org/title/cpi-detailed-report-58?browse=1970s.)

high enough to stop inflation from rising, however; and Volcker promptly raised the federal funds rate all the way to 17.6 percent by April 1980, using monetarist doctrine as cover. ("We don't control interest rates; we just try to make the money supply grow at a constant rate.") Yes, it was Volcker who conquered inflation. But the 630 basis points of tightening under Burns and Miller (from 4.6 percent to 10.9 percent) was certainly not nothing.

In total, the federal funds rate rose by a gigantic 1,300 basis points over the entire episode spanning January 1977 to April 1980. Approximately the last half of that tightening cycle came in just eight months under the determined inflation fighter, Paul Volcker. Unlike its vacillation in 1973–1974, the Federal Open Market Committee in 1979–1980 did not tarry over whether to deal with the new round of stagflationary shocks (truly bad luck) by easing or tightening monetary policy. It was tightening all the way. Volcker may or may not have been serious about following the monetarists' constant-money-growth doctrine, but he was certainly serious about disinflation—and he was *not* seeking a soft landing.

The landing itself was actually complex and came in two parts, each corresponding to its own recession. The National Bureau of Economic Research treats the short, sharp recession of 1980 as starting in January and ending in July, but it was mainly a one-quarter event (1980:II). Volcker's policy did not cause the 1980 recession. Carter's did.

The Carter White House had become convinced by early 1980 that excessive growth of consumer credit—rather than, say, inflationary expectations or excess money growth—was fueling inflation. Volcker and the Fed did not share this view. But Carter had the legal authority, left over from the Nixon years, to ask the Fed to impose credit controls, and he used it in March 1980. Looking back on that decision years later, Stuart Eizenstat (2018, p. 347), who was the chief domestic policy adviser in the White House at the time, termed it "a monumentally bad idea." It was.

Volcker and Harper (2018, p. 110) later recounted that the Fed viewed credit controls as "a transparently political ploy" at a time when "excessive credit wasn't the problem." But the politically astute Fed Chair felt that the central bank, though independent, could not just turn the president down flat. After all, Carter was supporting the Fed's tight-money policy despite evident political peril to himself. So Volcker asked the Fed staff to design credit controls that, for example, exempted borrowings to finance automobile and home purchases. Leaving a toothless tiger, right?

Apparently not. It turned out that Carter's well-publicized exhortations to eschew borrowing resonated strongly with an American public that was sick and tired of high inflation. According to Eizenstat (2018, p. 348), "People were so desperate to do their part to fight inflation, they tore up their credit cards as a patriotic act and sent the pieces to . . . the White House, accompanied by letters saying: 'Mr. President, we will cooperate.'" Consumer spending crumbled, taking GDP down with it. The second quarter of 1980 stands out as one of the worst quarters in post–World War II US history, with an annualized growth rate of *negative* 8 percent. The landing was painfully hard—but it was also blissfully short.

Episode 5: July 1980–January 1981

The slump was sharp enough, however, to induce Volcker and the Fed to back off from monetary tightening for a while. The Federal Open Market Committee reduced the federal funds rate from its peak of 17.6 percent in April 1980 all the way down to 9 percent by July 1980, even though double-digit inflation was not yet abating. It was not that the Volcker Fed suddenly decided to seek a soft landing. Rather, they became worried that the plane was headed for a catastrophic crash. That worry did not last long, however. The credit card mess passed, and the economy sprang back to life—growing at roughly an 8 percent annual rate in 1980:4 and 1981:1. The Committee promptly returned to raising interest rates rapidly, reaching a new peak federal funds rate of 19.1 percent by January 1981.

Thus, during the brief tightening cycle from July 1980 to January 1981, the Fed raised the federal funds rate by 1,005 basis points in just six months. Unsurprisingly, a severe 16-month recession followed from July 1981 to November 1982 (using the NBER dates). Quarterly GDP numbers show negative real growth in four of the six quarters spanning 1981:II through 1982:III, with a cumulative net decline of 2.1 percent. The unemployment rate shot up from about 7.5 percent in May–September 1981 all the way to 10.8 percent in November–December 1982, far eclipsing the 9 percent peak of 1975. The landing was hard indeed. The Fed, worried that it might get even worse, officially abandoned its monetarist experiment in October 1982, even though the core Consumer Price Index inflation rate was still about 6 percent.

But inflation was declining, with the usual lag. The core inflation rate was still in double digits (11.4 percent) when the Fed stopped raising rates in January 1981, and it was down to only 9.3 percent a year later. Seemingly scant progress. But then inflation fell rapidly, reaching 4.5 percent in December 1982 and 3 percent (a transitory low) by summer 1983. With the federal funds rate around 9 percent in both cases, the *real* federal funds rate was excruciatingly high. Paul Volcker is remembered as the man who broke the back of inflation. He did. And he did not believe he could do it with a soft landing.

Episode 6: February 1983–August 1984

It is questionable whether the next episode, a period of a year and a half, should be called a tightening at all. It was perhaps more an effort to recalibrate policy after the deep recession, combined with the unsuccessful experiment with money supply targeting, had left the federal funds rate too low. The facts are, however, that the Fed kept pushing the federal funds rate up for 18 consecutive months, accumulating to an increase of 313 basis points. The period from February 1983 to August 1984 certainly qualifies as a tightening by any objective standard, even though the Fed may not have been trying to push inflation down. During that tightening (and after), GDP growth was robust, unemployment tracked down steadily from its peak, and inflation was moderately stable in the 3–4.5 percent range.

If we look specifically at the "landing" and after, it was about as soft as you can imagine. After August 1984, GDP grew nicely for years, the unemployment rate fell slightly, and inflation drifted down to about 4 percent and stayed there. These were the early years of what came to be called the Great Moderation, a notable reduction in the variances of many macroeconomic variables which Stock and Watson (2003) attributed largely to good luck, though partially also to good monetary policy.

Episode 7: March 1988–April 1989

By early 1988, the unemployment rate had drifted down to 5.7 percent, which many economists viewed as a credible estimate of the natural rate of unemployment at the time; real GDP growth had averaged 4.5 percent over the previous year, and core inflation as measured by the Consumer Price Index had perked up slightly to 4.3 percent (versus 3.8 percent in the winter of 1986–1987). With the federal funds rate at 6.6 percent, the Federal Open Market Committee—now chaired by Alan Greenspan—decided to begin tightening, albeit ever so slowly. Over the ensuing 13 months, it raised the federal funds rate by 326 basis points (about 25 basis points per month) and then stopped.

History does not offer us redos, but I have long believed that the tightening cycle of 1988–1989 might have produced a perfect soft landing were it not for Saddam Hussein's invasion of Kuwait on August 2, 1990. Over the last half of 1989 and the first half of 1990, real GDP growth averaged 2.4 percent—below trend at the time, but well above zero. Then oil prices spiked; the price of a barrel of West Texas intermediate crude skyrocketed from around $16 in July to over $40 in October. While the 1990 oil shock was short (it was over by February 1991) it was not sweet. The violent spike in oil prices pushed the country into the short recession of 1990–1991, dashing any hopes of a soft landing. Bad luck for the Fed—and for the economy. GDP growth plummeted to an average annual rate of −1.7 percent over the three quarters spanning 1990:III–1991:I. The recession ended in March 1991, according to National Bureau of Economic Research dating, but the unemployment rate kept rising until the middle of 1992, leading the commentariat to dub the 1991–1992 period "the jobless recovery."

On the price front, core Consumer Price Index inflation rose from a low of 4.3 percent in the fall of 1989 to a high of 5.6 percent in early 1991. Headline inflation rose even more because of oil prices, topping 6 percent in late 1990. The Federal Open Market Committee mostly ignored this small inflationary uprising, however, perhaps recognizing its inherently temporary nature. It concentrated instead on helping the economy recover—just the opposite of what it had done when stagflation struck in 1979–1980. While the landing after this monetary tightening was on the hard side, the fault lies with Saddam Hussein, not Alan Greenspan.

Episode 8: December 1993–April 1995

The next tightening cycle produced the celebrated—and allegedly unique—perfect soft landing that helped make Alan Greenspan a central banking legend. Its start is usually dated as February 1994, when Greenspan made the Fed's first-ever public announcement that the Federal Open Market Committee had changed its target federal funds rate. Prior to that, market participants had to guess the Fed's decisions by watching the actions of the open-market desk as it bought or sold US Treasury securities. But as mentioned at the outset, I date the start as December 1993 instead because that was the low point for the *effective* federal funds rate.

At the time of that initial rate hike, core inflation in the Consumer Price Index was a quiescent 2.8 percent, and the unemployment rate was stable at 6.6 percent. This was then thought (erroneously) to be reasonably close to the "non-accelerating inflation rate of unemployment," or NAIRU—the unemployment rate that will not cause inflation to rise or fall.[9] Numbers like those did not obviously call for tighter monetary policy. But the federal funds rate had been sitting at 3 percent since September 1992, making the *real* federal funds rate about zero, and the Federal Open Market Committee was getting nervous about incipient inflationary pressures. It decided to act *preemptively* to head off inflation, a strategy that was applauded as innovative at the time. The Fed raised the target federal funds rate seven times in a year, including an eye-catching 75 basis point hike in November 1994—the only move that large in Greenspan's lengthy tenure as Fed Chair.

The results were fabulous. Inflation remained around 3 percent for two to three years and then drifted down slightly. The unemployment rate also trended down for most of the next *six* years, reaching a low of 3.8 percent in April 2000. Except for a short growth hiccup (down to a 1.3 percent annual rate) in the first half of 1995—which precipitated a small rate cut—real GDP growth rarely clocked in below 3 percent for the rest of the decade. There was certainly no recession.

So the landing was perfect. But notice that the Fed was *not* trying to push inflation down at the time and, in retrospect, it does *not* look like the economy had overshot full employment. Also notice that, unlike in 1990, the Fed's plans were not derailed by bad luck.

Episode 9: January 1999–July 2000

The late 1990s are remembered for strong growth, falling unemployment, stable and low inflation—and also for an amazing degree of Federal Reserve *forbearance* in the face of incoming data that might easily have provoked an inflation scare. Meyer (2004) labeled the monetary policy of this period Greenspan's "great call."

[9] A personal note: I was vice chair of the Federal Reserve Board during this period, and I was excoriated by many for suggesting, at the August 1994 Jackson Hole symposium, that the NAIRU might be 6 percent or lower (Blinder 1994).

Despite the exuberant boom, with real growth averaging above 4.5 percent over the four years 1996–1999 and unemployment falling, the Greenspan Fed did not flinch. It held the target federal funds rate at either 5.25 percent or 5.5 percent for about 2½ years (from February 1996 to September 1998) and then actually *lowered* it to 4.75 percent in three small cuts during the 1998 financial crisis. The first of these became known, with magnificent exaggeration, as "the 25 basis points that saved the world" because of its role in calming the roiling waters of international financial markets. At the time, many markets around the world had gone into shock after a Russian debt default and the collapse of the hedge fund Long-Term Capital Management in 1998.

Dating the start of the 1999–2000 tightening cycle is a bit tricky. The first *announced* hike in the *target* federal funds rate (from 4.75 percent to 5 percent) came in July 1999, but the *effective* federal funds rate had been drifting up for six months before that. For consistency with other dating decisions, I treat January 1999 as the trough month for the federal funds rate. By July 2000, when the effective rate peaked, it had risen by 191 basis points—a modest tightening. But helped along by the stock market crash of 2000, that was enough to precipitate a recession that began in April 2001. Real GDP growth was slightly negative in 2001:1 and 2001:3 (−1.3 percent and −1.6 percent annualized rates respectively), but the year 2001 as a whole nonetheless displayed modest positive growth (1 percent over 2000). Because the recession is so mild that it disappears in annual data, I have long called it the "recessionette." But call it what you will, the landing was pretty gentle.

As all this was happening on the real side, core inflation in the Consumer Price Index was flat at 2 percent during 1999, drifted up to 2.6 percent during 2000 and 2001, and then drifted back down to 1.9 percent during 2002. If the Fed was seeking Consumer Price Index inflation in the 1.5–2.0 percent range, as was widely believed at the time, it got there by late 2002.

All in all, while this episode does not qualify as a literal soft landing—there was, after all, a recession as judged by the National Bureau of Economic Research—it was certainly softish. The productivity surge of the late 1990s gave the Fed a dollop of good luck, and policymakers used it well.

Episode 10: May 2004–July 2006

The seeds of the Global Financial Crisis of 2007–2009 were sown by the excesses of the irrationally exuberant house-price boom of 2000–2006—and especially by Wall Street's reckless financial engineering with related paper assets. The Fed has been justly criticized for its disgracefully lax supervisory policies, a failing it shared with the other financial regulatory agencies. Thus, there may have been a different kind of soft/hard lesson here: Supervisors who are too soft on banks' risky lending practices for too long risk a hard landing when bubbles burst. The landing that started in late 2007 was certainly very hard, earning the name given it at the time: the Great Recession.

But this paper is about *monetary* policy, not *regulatory* policy. Was the Fed's allegedly lax monetary policy in 2003–2004, when it held the federal funds rate at either 1 percent or 1.25 percent for almost two years, a major cause of the bubble and, by inference, of the crash that followed—as argued, for example, by Taylor (2009)? I think not, for several reasons.

First, if the Fed was late to start raising interest rates, it was not *very* late. Precious few observers in 2003 were yelling "housing bubble"—not even the celebrated bubble-finder, Robert Shiller (Case and Shiller 2003; for more examples, see Blinder 2013, pp. 32–35). The Fed's first official rate hike, of 25 basis points, occurred in June 2004, and that was followed by 16 more such hikes for a total of 425 basis points. The effective federal funds rate matched this tightening cycle almost exactly.

Second, the Fed had good reason to be late: The recovery from the 2001 recession was not just *jobless*, as in 1991, it was actually plagued by further job *losses*. The NBER recession trough came in November 2001, but net job losses continued through August *2003*. The unemployment rate, which was 4.3 percent when the recession began in March 2001, did not return to the 4's again until the end of *2005*. For a central bank with a dual mandate, those numbers were worrisome. Besides, inflation was moribund or drifting down for most of the time. So why tighten?

Third, the 425 basis points of Fed rate hikes that did occur proved insufficient to burst the house-price bubble. How much more, then, would it have taken to do the job? No one can possibly know, but it seems a plausible guess that much tighter monetary policy might have precipitated a recession before it burst the house-price bubble.

All that said, the eventual "landing" was undoubtedly very hard. The recession that the NBER says began in January 2008 lasted six quarters and pulled real GDP down by a cumulative 3.8 percent. The operative question for this paper is: Did tight money cause the Great Recession?

I am dubious in the extreme. First, notice the long time lag. The Fed's tightening cycle began in June 2004 and *ended* in June 2006. The first negative quarter of real GDP growth was not until 2008:I, and the really severe declines did not occur until 2008:IV (−8.5 percent) and 2009:I (−4.6 percent). That is a very long lag between cause and effect. Indeed, just days before the September 2008 Lehman Brothers bankruptcy, the Congressional Budget Office (2008) forecast growth rates of 0.9 percent and 1.8 percent over the four quarters of 2008 and 2009, respectively. Second, we know that much stronger forces than the Fed's tightening in 2004–2006 were buffeting the economy in 2008, especially after the Lehman bankruptcy. The entire financial system was teetering on the brink of collapse, leading the Fed to cut interest rates to the bone.

So to my mind, the deep recession of 2007–2009 was *sui generis*. It was surely "hard," but it was not a "landing" from an episode of tight money. I can even imagine that the US economy might have experienced a soft landing had the financial system not imploded. As it was, real growth averaged a modest 2.2 percent in

2007, and dropped to 0.3 percent in the first half of 2008—still positive until the Lehman shock in September 2008, after which everything fell apart.

Episode 11: November 2015–January 2019

Given the length of time covered by this final Fed tightening—more than three years—one might not want to call this a tightening cycle at all. It was more like a long normalization of interest rates after seven years of a "zero" interest policy.[10] Nonetheless, the effective federal funds rate did rise by 228 basis points during this time and a recession followed, albeit a very short and deep one clearly caused by the COVID pandemic. The episode thus meets the objective criteria for a monetary policy tightening. But while the landing was monumentally hard, the Fed's tightening was obviously not the cause.

The Scoreboard

What, then, do we conclude about the Fed's demonstrated ability to land the economy softly when it fights inflation? Table 1 is a short tabular summary of the history recounted in this paper. The glass can admittedly be seen as either half full or half empty. I read it as more than half full, however, and here is why.

The final two episodes of monetary tightening (2004–2006 and 2015–2019) ended painfully, for sure. But it is abundantly clear that these two deep recessions were *not* the products of tight money. That leaves nine tightenings to consider. Of these, the three episodes in the 1970s and early 1980s—when the Great Inflation peaked and then dissipated—were undoubtedly followed by seriously hard landings. But in two of these three cases (1977–1980 and 1980–1981), that was clearly a policy choice: the Volcker Fed was not trying for a soft landing. Double-digit inflation had to be conquered, even at high cost. In the other 1970s case (1972–1974), a recession was probably inevitable whatever the Fed did, given the severity of the supply shocks buffeting the economy. Monumentally bad luck.

That leaves six cases, and in all but one of them I would characterize the landing as on the soft side. The lone exception, the year-long tightening cycle of 1988–1989, was followed by the 1990–1991 recession. But as I have argued, it might have been a perfect soft landing were it not for the first Gulf War in 1990. The Greenspan Fed was trying for soft, and they almost succeeded, but bad luck intervened. So it appears that the Fed's reputation for causing hard landings with tight money derives mainly from conquering the 1970s inflation—which took three landings.

[10] The Fed has steadfastly refused to go literally to a policy interest rate of zero percent, much less beyond. Its "zero" interest rate policy has always meant posting a range between zero and 25 basis points. During the seven years from December 2008 to December 2015, the lowest effective funds rate (monthly average) was 7 basis points.

Table 1
The Eleven Monetary Policy Tightenings since 1965

	Dates	*Total basis points (effective rate)*	*Inflation two years later*	*Landing hard or soft?*
1	Sept. 1965–Nov. 1966	174	Higher	Quite soft—but was it a landing?
2	July 1967–Aug. 1969	540	Lower	Softish
3	Feb. 1972–July 1974	962	Higher	Hard
4	Jan. 1977–Apr. 1980	1,300	Lower	Hard
5	July 1980–Jan. 1981	1,005	Lower	Hard
6	Feb. 1983–Aug. 1984	313	Lower	Very soft—but was it a landing?
7	Mar. 1988–Apr. 1989	326	Higher	Likely would have been a soft landing
8	Dec. 1993–Apr. 1995	309	Lower	Soft
9	Jan. 1999–July 2000	191	Same	Softish
10	May 2004–July 2006	424	Higher	Hard—but not due to Fed
11	Nov. 2015–Jan. 2019	228	Lower	Hard—but not due to Fed

That said, there is also a plausible argument for the "half empty" view. After all, leaving aside the last two cases once again, six of the remaining nine tightenings were followed by recessions. The 1965–1966 tightening was a mere 174 basis points, perhaps too little to be considered a serious tightening. The 1983–1984 episode might be more accurately seen as a post-monetarist adjustment back to interest-rate targeting than as a policy tightening intended to slow the economy.

Readers may take their choice. I have expressed mine, which is this: If the need to fight inflation is not too extreme, and serious adverse events like wars or supply shocks do not intervene, the Federal Reserve has shown itself capable of engineering a landing that either does not induce a recession or, if it does, induces a small one. It has done so several times. Furthermore, its dual mandate pushes it to attempt the feat.

On the other hand, steering the economy to a soft landing is a delicate task which can be upset by, among other things, external shocks—external to monetary policy, that is. Historically, oil shocks and food shocks are the prime examples. But a big fiscal shock like spending on the Vietnam War or the Reagan tax cuts might also do the trick. Jimmy Carter's credit controls are another example.

Perhaps most obviously, the likelihood of landing the economy softly depends on how high the "plane" was flying before it began to descend. Alan Greenspan's perfect soft landing in 1994–1995 did not bring inflation down at all; it merely avoided what was thought to be a potential rise in the inflation rate. Paul Volcker, in stark contrast, inherited a double-digit inflation rate in 1979, which he brought down to about 4 percent. It was a long way down, and the landing was rough.

Since March 2022, the Fed has been tightening monetary policy once again—raising interest rates to fight the worst inflation since the early 1980s. Episode 12 is not over yet, and I cannot predict how high interest rates will go, how long it will take to beat inflation, nor how hard or soft the eventual landing will be. What is clear, however, is that, between the COVID-induced supply disruptions, the oil shock, and the food shock, the luck factor has run strongly against them. To achieve another soft landing under these circumstances, the Fed will have to be skillful indeed.

References

Abrams, Burton A. 2006. "How Richard Nixon Pressured Arthur Burns: Evidence from the Nixon Tapes." *Journal of Economic Perspectives* 20 (4): 177–88.

Bernanke, Ben S. and Alan S. Blinder. 1992. "The Federal Funds Rate and the Channels of Monetary Transmission." *American Economic Review* 82 (4): 901–21.

Blinder, Alan S. 1994. "Overview." In *Proceedings of the Federal Reserve Bank of Kansas City Symposium on Reducing Unemployment: Current Issues and Policy Options*, 329–42. Jackson Hole, WY: Federal Reserve Bank of Kansas City.

Blinder, Alan S. 2013. *After the Music Stopped: The Financial Crisis, the Response, and the Work Ahead.* New York: Penguin.

Blinder, Alan S. 2022. *A Monetary and Fiscal History of the United States, 1961–2021.* Princeton, NJ: Princeton University Press.

Blinder, Alan S., and William J. Newton. 1981. "The 1971–1974 Controls Program and the Price Level: An Econometric Post-Mortem." *Journal of Monetary Economics* 8 (1): 1–23.

Blinder, Alan S., and Jeremy B. Rudd. 2013. "The Supply-Shock Explanation of the Great Stagflation Revisited." In *The Great Inflation: The Rebirth of Modern Central Banking*, edited by Michael D. Bordo and Athanasios Orphanides, 119–75. Chicago: University of Chicago Press.

Burger, Albert E. 1969. "A Historical Analysis of the Credit Crunch of 1966." *Federal Reserve Bank of St. Louis Review* 51 (9): 13–30.

Burns, Arthur F. 1979. "The Anguish of Central Banking." Paper presented at the Per Jacobssen Lecture, Belgrade, September 30.

Case, Karl E., and Robert J. Shiller. 2003. "Is There a Bubble in the Housing Market?" *Brookings Papers on Economic Activity* 33 (2): 299–362.

Congressional Budget Office (CBO). 2008. *The Budget and Economic Outlook: An Update*, Washington, DC: United States Congressional Budget Office.

Council of Economic Advisers (CEA). 1968. *Economic Report of the President.* Washington, DC: US Government Printing Office.

Dornbusch, Rüdiger. 1997. "Why This Recovery Won't Fall Off the Track Soon." Bloomberg, August 11. https://www.bloomberg.com/news/articles/1997-08-10/why-this-recovery-wont-fall-off-the-track-soon.

Eizenstat, Stuart E. 2018. *President Carter: The White House Years.* New York: St. Martin's Press.

Federal Reserve. 2012. "Federal Reserve issues FOMC statement of longer-run goals and policy strategy." January 25, https://www.federalreserve.gov/newsevents/pressreleases/monetary20120125c.htm.

Federal Reserve. 2020. "2020 Statement on Longer-Run Goals and Monetary Policy Strategy." August 27, https://www.federalreserve.gov/monetarypolicy/review-of-monetary-policy-strategy-tools-and-communications-statement-on-longer-run-goals-monetary-policy-strategy.htm.

Friedman, Milton. 1961. "The Lag in Effect of Monetary Policy." *Journal of Political Economy* 69 (5): 447–66.

Heller, Walter W. 1966. *New Dimensions of Political Economy.* Cambridge, MA: Harvard University Press.

Holston, Kathryn, Thomas Laubach, and John C. Williams. 2017. "Measuring the Natural Rate of Interest: International Trends and Determinants." *Journal of International Economics* 108 (S): S59–75.

Johnson, Noah. 2017. "A Comparison of PCE and CPI: Methodological Differences in U.S. Inflation Calculation and Their Implications," US Bureau of Labor Statistics, November. https://www.bls.gov/osmr/research-papers/2017/st170010.htm.

Meltzer, Allan H. 2009. *A History of the Federal Reserve.* Vol. 2, Bk. 2, 1970–1986. Chicago: University of Chicago Press.

Meyer, Laurence H. 2004. *A Term at the Fed: An Insider's View.* New York: Harper Collins.

Okun, Arthur M. 1970. *The Political Economy of Prosperity.* Washington, DC: Brookings Institution Press.

Reis, Ricardo. 2021. "Losing the Inflation Anchor." *Brookings Papers on Economic Activity* 51 (2): 307–61.

Shapiro, Adam H. and Daniel J. Wilson. 2022. "Taking the Fed at Its Word: A New Approach to Estimating Central Bank Objectives Using Text Analysis." *Review of Economic Studies 89 (5): 2768–2805.*

Stock, James H., and Mark W. Watson. 2003. "Has the Business Cycle Changed? Evidence and Explanations." In *Proceedings of the Federal Reserve Bank of Kansas City Symposium on Monetary Policy and Uncertainty: Adapting to a Changing Economy*, 9–56, Jackson Hole, WY: Federal Reserve Bank of Kansas.

Taylor, John B. 2009. *Getting Off Track: How Government Actions and Interventions Caused, Prolonged, and Worsened the Financial Crisis.* Stanford, CA: Hoover Institution Press.

Volcker, Paul A., and Christine Harper. 2018. *Keeping at It: The Quest for Sound Money and Good Government.* New York: PublicAffairs.

Journal of Economic Perspectives—Volume 37, Number 1—Winter 2023—Pages 121–144

Monetary Policy and Inequality

Alisdair McKay and Christian K. Wolf

Monetary policy affects incomes, employment rates, asset prices, and interest rates. Because different households hold different financial assets, work in different sectors, and are differentially attached to the labor force, it follows that monetary policy will almost certainly have heterogeneous effects across households. The last decade has seen an explosion of empirical and theoretical research on the links between monetary policy and inequality. In this article, we take stock of this important research agenda. We organize our discussion around three main questions.

First, how unevenly distributed are the effects of monetary policy, and why? Monetary stimulus—that is, lower nominal interest rates—affects households through many different channels. Most directly, households may pay lower interest rates on their debts and are likely to earn lower returns on their savings. In addition, changes in interest rates also set into motion changes in the broader economy that indirectly affect households: jobs become easier to find, wages and prices increase, and asset prices rise. Taking any one of these channels in isolation, the effects of a monetary policy easing can look very uneven. For example, many households will not benefit at all from an increase in stock prices, simply because they do not own

■ *Alisdair McKay is a Senior Research Economist, Federal Reserve Bank of Minneapolis, Minneapolis, Minnesota. Christian K. Wolf is Assistant Professor of Economics, Massachusetts Institute of Technology, and a Faculty Research Fellow, National Bureau of Economic Research, both in Cambridge, Massachusetts. Their email addresses are alisdair.mckay@mpls.frb.org and ckwolf@mit.edu. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.1.121.

any stocks. When aggregating across all of these transmission channels, however, the gains appear much more evenly distributed. Roughly speaking, low-income households benefit from a tighter labor market, middle-class households benefit from lower mortgage rates, and wealthy households benefit from capital gains on assets. To reach this conclusion, we review recent empirical work on the incidence of monetary policy across households. We also provide descriptive measures of the heterogeneity in household balance sheets and their exposure to changes in monetary policy.

Second, does a more careful account of the microeconomic effects of monetary policy affect our understanding of its propagation to the macroeconomy at large? Yes and no. Theoretical analyses that dig into the microeconomic propagation of monetary policy have materially altered our views on transmission channels. For example, we have learned that an important component of the transmission of monetary policy to consumer spending is actually indirect: by changing nominal interest rates, monetary policy first of all directly increases consumer and firm demand; this increase in demand then in general equilibrium leads to higher income, which in turn leads to meaningful second-round effects on consumer spending. At the same time, however, there is rather limited scope for such microlevel studies to change our overall view on the macro effects of monetary policy. The reason is simple: we already have good empirical evidence on the overall response of aggregate output and inflation to changes in interest rates, and any structural model of monetary policy transmission—with or without microeconomic heterogeneity—needs to be broadly consistent with that evidence. That said, the more detailed view of the micro-level effects of monetary policy that emerges from recent research does suggest new reasons to expect the economy's sensitivity to monetary policy to vary over time and with the state of the business cycle.

Third, how—if at all—should the interaction between monetary policy and inequality affect the behavior of central banks? We argue that even if the central bank's mandate includes distributional concerns, appropriate policy is unlikely to differ too much from the optimal policy of a central bank that is solely focused on macroeconomic goals like stabilizing inflation and aggregate activity. The reason is related to our answer to the first question: because the effects of monetary policy are relatively evenly distributed, the scope for achieving distributional objectives through monetary policy is likely to be rather limited.

The Distributional Effects of Monetary Policy

Our objective in this section is to assess how monetary policy affects consumption across different groups of households. One could, in principle, instead ask how monetary policy affects other measures of inequality, such as income or wealth. We will ask these questions as well, but with the end goal of translating income and wealth into consumption. We do so because, while of course consumption is not

synonymous with utility, it is more directly related to a household's well-being than are its wealth and income.

At any given point in time, monetary policy will of course explain very little of the overall differences in consumption across different groups of households. Therefore, we discuss how consumption *changes* across household groups following a *change* in monetary policy. These consumption responses will depend—at least to some degree—on the institutions of the economy (for example, whether mortgages are mostly fixed-rate or floating-rate). We focus primarily on the US context, though many of the forces we describe are likely to be similar across countries. We will also phrase our discussion in terms of the effects of monetary stimulus; in many respects a monetary contraction would have the mirror image effects, though there are some important sources of asymmetry that we will note. We proceed in two steps. First, we review the key channels through which monetary policy will affect consumption. Second, we combine all of those channels to arrive at overall conclusions for how monetary policy will affect the consumption of heterogeneous households.

Channels of Transmission

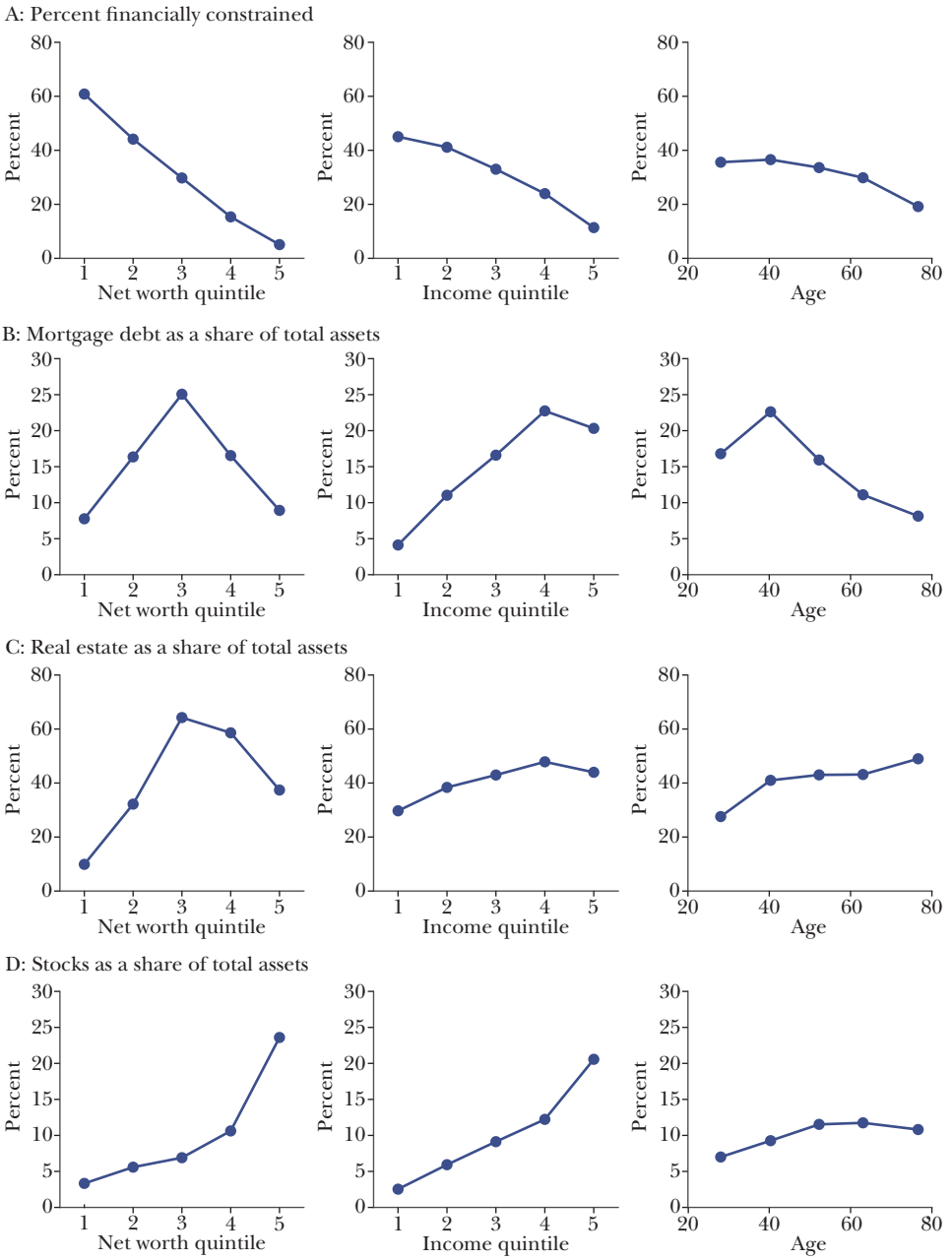
Monetary policy affects the consumption of an individual household by changing the prices, wages, interest rates, and opportunities it faces. Many of these changes are indirect in nature; for example, expansionary policy may lead to a tighter labor market, thus resulting in higher wages for workers. This section considers some of the main channels of transmission from monetary policy decisions to household consumption decisions: via income, revaluation of nominal contracts, mortgages, asset prices, and intertemporal substitution.[1] For each channel, we will describe its strength in the aggregate as well as its heterogeneous incidence across households. To do so, we will combine empirical evidence on the aggregate effects of changes in monetary policy with data on heterogeneity in household finances. That data is taken from the 2019 Survey of Consumer Finance (Federal Reserve Board 2019), a nationally representative survey of households that collects data on their income, assets, and liabilities.

Throughout this section we will repeatedly refer to Figure 1. Each panel of three figures (the rows of the figure) shows a different feature of household balance sheets. The left column of the figure splits households into five wealth quintiles and reports the average within each quintile. The middle and right columns do the same splitting households into income and age quintiles, respectively. We focus on these dimensions of heterogeneity across households as they interact most closely with the mechanisms studied in the recent literature.

Income. Expansionary monetary policy stimulates the aggregate economy and thereby raises labor income. These gains are unlikely to be distributed equally, as labor incomes of low-income households tend to be disproportionately exposed to the business cycle (Okun 1973; Guvenen, Ozkan, and Song 2014; Guvenen et al.

[1] A theoretical decomposition of individual consumption responses to monetary policy that includes these channels can be found in Auclert (2019).

Figure 1
Some Summary Measures of Household Balance Sheets



Source: Data from the 2019 Survey of Consumer Finances. Author's calculations shown in online Appendix.
Note: Each figure divides the sample into quintiles by net worth, income, or age and then plots the mean level within that quintile. Panel A shows the fraction of households that have liquid assets of less than two weeks' income. Panel B shows the mean ratio of mortgage debt to total assets. Panel C shows the share of real estate in household asset holdings. Panel D shows the share of stocks in households' assets.

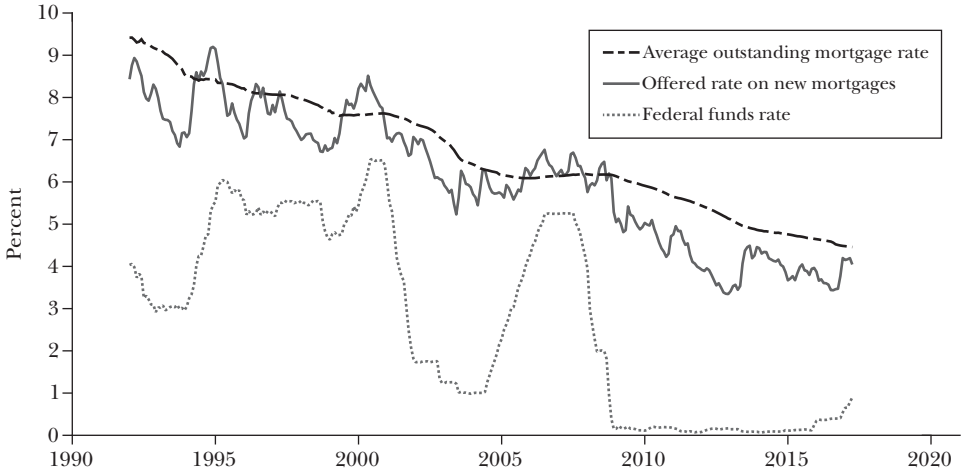
2017; Patterson 2022). This pattern holds both for business cycles in general, and also for expansions and contractions that result from changes in monetary policy in particular (Andersen et al. 2021; Amberg et al. 2021; Holm, Paul, and Tischbirek 2021). The inequality in earnings resulting from this income incidence channel is quite pronounced. For example, data from the recession of 1979–1983 (which was arguably caused by a monetary contraction) shows that the earnings losses for low-income households were many times larger than those for high-income households (Guvenen, Ozkan, and Song 2014).

These heterogeneous responses of income are then further amplified when translated to changes in consumption. In particular, low-income households are more likely to be financially constrained—that is, with no savings and no access to credit. Without a buffer of savings or credit, such households are then likely to have a stronger response of consumption to a change in income. To illustrate the empirical relevance of this point, we report in panel A of Figure 1 the fraction of households that are financially constrained. To construct the figure, we follow Kaplan, Violante, and Weidner (2014) and classify households as financially constrained if their liquid assets amount to less than two weeks' worth of income. According to this definition, it is possible even for high-net-worth households to be financially constrained if their assets are held in illiquid forms (such as real estate). Overall, we classify 31 percent of households as being constrained in this fashion, with households of lower net worth and with lower income much more likely to be constrained.

Revaluation of nominal contracts. Expansionary monetary policy raises the general level of prices by temporarily raising the rate of inflation. In most borrowing and lending arrangements, the contracting parties agree to a repayment that is set in nominal terms, and so a surprise increase in the price level will reduce the real value of the repayment. Doepke and Schneider (2006) document the heterogeneity in household exposure to such surprise inflation. A typical middle-class household has substantial nominal debts in the forms of mortgages, auto loans, credit card debts, and student loans. For a typical household, nominal assets (like bank deposits and bond holdings) are smaller than nominal debts, and so it follows that a surprise increase in inflation will lower the real value of their debts by more than it will lower the real value of their assets. This is especially true of young households—a group that tends to have large mortgage balances. Old, rich households, on the other hand, tend to have more nominal assets than nominal liabilities and so the net worth of these households declines after an unexpected increase in the price level.

How large are these effects? It is widely believed that inflation is actually fairly insensitive to short-run changes in monetary policy (for example, Mavroeidis, Plagborg-Møller, and Stock 2014). On the other hand, since most central banks nowadays try to stabilize rates of inflation, even small changes in the inflation rate can accumulate over time to have a substantial effect on the overall price level. In particular, as we describe further below, a change in monetary policy that raises GDP by 1 percentage point for two years approximately leads to a 0.5 percentage point permanent increase in the price level. For a household with a mortgage

Figure 2
Monetary Policy and Mortgage Interest Rates



Source: Berger et al. (2021), Freddie Mac, and Federal Reserve Board.
Note: The outstanding rate is the average rate currently paid by homeowners. The federal funds rate is an overnight interest rate and in particular the target rate for monetary policy implementation with data from the Federal Reserve. The offered rate is the average rate currently offered by mortgage lenders.

balance that is twice its annual income (about the average ratio for young homeowners), this channel represents a decrease in the real value of their liabilities equal to 1 percent of their annual income.

Mortgages. A homeowner with a mortgage not only benefits from the change in the price level, but may also benefit from a reduction in the nominal interest rate. Mortgage interest rates are tightly linked to monetary policy: a 1 percentage point change in the federal funds rate typically translates to roughly a 0.5 percentage point change in the 30-year mortgage rate (for a review of evidence, see Wong 2021 and references therein). The design of mortgage contracts varies across countries. In the United States, most mortgages have fixed nominal interest rates, but the borrower is free to repay the loan at the time of their choosing. This option to repay the loan creates an asymmetry: when nominal interest rates fall, homeowners can benefit by refinancing their loans at lower interest rates; if rates rise, homeowners can simply keep their original loans. Figure 2 plots the average interest rate paid on outstanding mortgages (as computed by Berger et al. 2021) along with the rate currently offered on new 30-year mortgages, as well as the federal funds rate. Offered mortgage rates, which are long-term rates, are less volatile than the federal funds rate. Moreover, as most borrowers have fixed-rate contracts and refinancing is gradual, the rates they actually pay are even smoother than the offered rates. Finally, due to the asymmetry mentioned above, average mortgage rates tend to co-move more closely with the federal funds rate following rate cuts than rate hikes.

Panel B of Figure 1 shows what kind of households will tend to gain from lower interest rates through this mortgage channel. In particular, the figure shows the ratio of mortgage debt to household assets. We see that mortgage debt is particularly important for households in the middle of the wealth distribution. Moreover, the ratio of mortgage debt to assets tends to decline with age, so young households are more exposed to changes in mortgage rates than are older households. Returning to our example of a young household with a mortgage balance of twice their income, a 1 percentage point reduction in their mortgage rate implies a 2 percent increase in disposable income for the life of the loan, which is typically between five and ten years.

Of course, whenever households pay lower interest rates on their debts, somebody else is receiving less interest income. In the US mortgage market, the lenders are often the owners of mortgage-backed securities—financial contracts that entitle the owner to receive the principal and interest payments on a pooled group of mortgages. It is generally difficult to say how this lost interest income will feed back to households, as mortgage-backed securities are typically held by financial institutions, governments, and foreign investors rather than outright by households.[2] Because this loss of interest income is directly borne by financial institutions, governments, and foreign investors, it is rather unlikely to have a substantial effect on household consumption decisions, at least in the short run.

Asset prices. Expansionary monetary policy tends to increase the values of long-lived assets such as stocks and real estate. As real interest rates decline, the present discounted value of future cash flows increases, thus leading to a revaluation of assets and liabilities. At first glance, this channel may appear to be a key channel of monetary policy-induced redistribution: asset-holders get wealthier when rates fall. However, lower interest rates also mean lower expected returns on these assets—a force pushing in the opposite direction. Which of these effects dominates depends on the horizon at which the asset-holder plans to consume.[3] Fagereng et al. (2022) provide an intuitive way for thinking about the redistributive effects of changes in asset prices: those who plan to sell the asset benefit when its price increases, and those who plan to buy the asset are harmed. Additional important effects relate to the role of assets as collateral and as buffers against changes in income: through these channels, higher asset prices may lead households to spend more even if they do not plan to buy or sell the asset. In what follows, we elaborate on the connection between changes in wealth and changes in consumption for two important asset classes: stocks and housing.

Stock prices are highly sensitive to changes in interest rates (Bernanke and Kuttner 2005; Bauer and Swanson 2022). As discussed above, the effect of stock

[2] Data from 2010 shows that the US government and foreign investors owned nearly 50 percent of outstanding agency mortgage-backed securities (Tracy and Wright 2012).

[3] More precisely, what matters is how a household's plan to consume at various dates lines up with the existing claims to cash flows across those dates. Households with front-loaded consumption and back-loaded cash flows benefit from lower rates (Auclert 2019; Greenwald et al. 2021; Fagereng et al. 2022).

price changes on household consumption is likely to depend on the future savings and consumption plans of the household. Empirical evidence overall suggests that higher stock market wealth does translate to an immediate (if moderate) increase in consumer spending, with an extra $1 of stock wealth increasing consumption by $0.03 (Chodorow-Reich, Nenov, and Simsek 2021). Across households, the ownership of stocks is highly concentrated, with wealthy households holding the vast majority. Panel D of Figure 1 reveals that wealthy households also devote a much larger share of their portfolios to stocks.

House prices tend to increase following expansionary monetary policy (Iacoviello 2005). Moreover, housing is particularly important as a share of household balance sheets for the middle class (for example, see panel C of Figure 1).[4] Many homeowners expect to remain in their houses for many years; since they then do not plan to either buy or sell the asset, it may seem that they are unaffected by a change in house prices. Intuitively, following a monetary easing, those households now own a more valuable house, but they now also want to live in a more expensive house. However, houses often serve as valuable sources of liquidity for households, either as collateral for loans or through the reassurance that they can rely on their home equity as a financial backstop in the future (for example, Berger et al. 2018). Empirical evidence suggests that homeowners as a group overall do increase their consumption when their homes appreciate—a "housing wealth effect." In particular, recent estimates show that an additional dollar of housing wealth leads to an increase in consumption of between $0.03 (Guren et al. 2021) and $0.07 (Mian, Rao, and Sufi 2013).

Intertemporal substitution. Finally, monetary policy can induce households to substitute consumption across time. Intuitively, the real interest rate is the price of consumption today relative to consumption in the future, and a decline in this relative price should result in households increasing consumption today and reducing consumption in the future. How strongly households respond to these incentives depends on the types of consumption we are considering. For nondurable goods and service consumption, empirical evidence (Best et al. 2020) suggests that this intertemporal substitution effect is quite small; it is also homogeneous across households, thus limiting its redistributive effects.

Assessing Consumption Effects through Direct Measurement

We have seen that monetary policy affects household balance sheets through several distinct margins, with each channel likely to benefit different groups of households. Ultimately, we are interested in how the sum of these changes translates to household consumption. In the rest of this section, we describe two broad approaches to answering this question: a direct approach (this subsection) and an indirect approach (the next one).

[4] Here we study asset price changes and thus we focus on homeowners rather than renters. Renters that do not plan to purchase a home are not directly affected by these asset price changes, though they could be affected indirectly if rents adjust to reflect these changes in home prices.

A direct measurement approach, in which one measures consumption at the household level and then asks how the distribution of consumption responds to changes in monetary policy, is perhaps the most natural way to proceed. The key challenge with this approach is that high-quality data on individual consumption is needed—the data need to allow the researcher to tease out the role of monetary policy among the many other (and actually more important) factors that also affect inequality across households.

US data. For the US economy, the best source for individual consumption data is arguably the Consumer Expenditure Survey—a nationally representative survey of households that is conducted quarterly by the Census Bureau on behalf of the Bureau of Labor Statistics. Coibion et al. (2017) use these data to create quarterly time series of statistics that summarize the consumption distribution—for example, the tenth, fiftieth, and ninetieth percentiles of the consumption distribution in each quarter. They then explore how these distributional measures evolve after a plausibly exogenous change in monetary policy. They find that expansionary monetary policy reduces consumption inequality, because the ninetieth percentile of the distribution moves closer to the tenth and fiftieth percentiles. Chang and Schorfheide (2022) also use Consumer Expenditure Survey data for the same purpose, though with a somewhat different econometric approach and a different measure of monetary policy. They instead find that expansionary policy *increases* consumption inequality, with the top-end of the distribution now moving away from the rest. Overall, one of the challenges with this direct approach—and a potential reason behind the conflicting findings in prior work—is that the contribution of monetary policy shocks is small relative to other factors that affect a cross-section of households. As a result, sampling variation can be an important obstacle in isolating the heterogenous effects of policy.

One way of circumventing these challenges is to estimate heterogeneous consumption effects of monetary policy across broader groups of households, notably across homeownership status. Cloyne, Ferreira, and Surico (2020) find that the consumption levels of homeowners with a mortgage and renters respond to changes in interest rates by similar percentage amounts, while the consumption of homeowners without a mortgage does not appear to react as strongly. Specifically, they find that a 1 percentage point reduction in nominal interest rates leads to about a 1 percentage point increase in nondurable consumption for mortgagors and renters, and no statistically significant spending response for homeowners without mortgages. This broad gradient by homeownership is then likely to translate into heterogeneity in the effects of monetary policy by wealth, income, and age. First, for the United States, the fraction of households in the Survey of Consumer Finance that has a mortgage is about 60 percent for the top three wealth quintiles and rises steadily across the income quintiles, reaching 73 percent for the highest quintile. This pattern suggests consumption gains that are somewhat increasing in wealth and income. Second, older households are less likely to have a mortgage, so younger households should benefit by more than older households. The implied gradient in consumption responses by age is consistent with results reported in Wong (2021). She finds that the consumption response to monetary policy is concentrated among

homeowners who take out a new mortgage (either to refinance an existing one or purchase a new home) after the change in monetary policy.

Evidence from other countries. Another way of circumventing the challenge of sampling variation is to use an administrative dataset that contains the entire population of households. Such data is not available for the United States. However, Holm, Paul, and Tischbirek (2021) construct measures of household consumption using Norwegian tax data on the incomes and assets of households. The key insight behind this approach is that household-level consumption can be imputed reasonably well by using the assets held by a household at the start of the year, adding the income received during that year, and then subtracting the assets held at the end of the year. A further key benefit of the Norwegian data is its panel structure, allowing Holm et al. to measure the *change* in consumption at the household level following a change in monetary policy.

The results of Holm, Paul, and Tischbirek (2021) suggest that expansionary monetary policy has *U*-shaped effects on consumption across the wealth distribution, with asset-poor and asset-rich households increasing their consumption somewhat more than households in the middle. The authors find similar patterns for disposable income, reflecting strong responses of nonfinancial income at the bottom of the asset distribution and financial income at the top of the distribution. The largest changes in consumption and disposable income occur with a substantial delay after the change in policy. At these horizons, the least-wealthy and wealthiest groups increase consumption by 1.5 to 2 percentage points, while those in the middle increase consumption by about 1 percentage point. Overall, we view the Holm, Paul, and Tischbirek (2021) study as particularly informative given its use of high-quality panel data. However, caution should be used in applying the Norwegian results in a US context; for example, Norwegian mortgages usually have adjustable interest rates, and Norwegian households have relatively few direct holdings of stocks.

Assessing Consumption Effects through Indirect Measurement

An indirect approach offers an alternative to direct measurement: by using a combination of theory and empirical evidence, one can aggregate the various individual channels of monetary transmission discussed earlier into a total effect on household consumption.[5] On the whole, taking into account all channels, we will argue that the consumption changes from a monetary easing appear relatively evenly distributed in the cross-section of households.

Ingredients. Table 1 presents the ingredients that we use for our indirect calculation of household consumption responses to a monetary easing. The table lists a variety of "prices" to which households are exposed—that is, the transmission channels discussed earlier and how sensitive those prices are to changes in monetary

[5] Auclert (2019) is an important and well-known example of the indirect approach to assessing the distributional effects of monetary policy. Slacalek, Tristani, and Violante (2020) focus on European data and pursue an approach closely related to the one we present here.

Table 1
Inputs for Calculation of Consumption Effects of Monetary Policy

		Marginal prop. to consume	
	Price change	*Unconstrained*	*Constrained*
Labor earnings	1.3%	0.05	0.5
Business income	1.0%	0.05	0.5
Interest income	–1.5%	0.05	0.5
Return on stocks	4.8%	0.03	0.03
Return on housing	0.6%	0.03	0.03
Return on cash	–0.56%	0.05	0.5
Mortgage rates	–0.65%	0.8	1.0
Other interest rates	–0.88%	0.8	1.0

Source: See online Appendix for details.
Note: Each row of the table corresponds to a channel of monetary policy effects on household consumption. The table lists how strongly the price or income associated with that channel responds to monetary policy and how strongly households spend out of that income category. The constrained column applies to households with liquid assets less than two weeks' worth of income. Changes in incomes, assets, returns, and interest rates are in real terms.

policy. To construct the values in the second column, we estimate how those various prices respond to a plausibly exogenous change in interest rates induced by monetary policy (with details on identification and estimation presented in Appendix A). All estimates are in real terms and have been scaled to correspond to a monetary stimulus that leads real GDP to increase by 1 percent on average over the first two years following the change in policy. The first row shows that labor earnings respond slightly more than GDP, while the second row reveals that business income moves about one-for-one with GDP. Further down the table, we see that stock prices are very sensitive to monetary policy, increasing by about five times more than GDP. Lower real interest rates—including in particular lower mortgage rates—reflect both lower nominal rates as well as an increase in the price level.

The third and fourth columns of the table list an assumed marginal propensity to consume for each type of income—that is, the strength of the consumption change following a change in income, for each income category. For example, a value of 0.05 means that a $1 increase in income would lead to an increase of $0.05 in consumption. It is important to note that a household's marginal propensity to consume is likely to depend on its (financial) circumstances; for example, economic theory suggests that a household that is financially constrained may spend strongly out of any additional income, while households with access to savings or credit are more likely to save additional income or use it to pay down debt. To capture these effects in a transparent way, we assume that households with few liquid assets are financially constrained and have a high marginal propensity to consume out of transitory changes in income; high liquid-wealth, unconstrained households, on the other hand, will have lower marginal propensities to consume. The third and fourth

columns of the table reflect this split—one for financially constrained households and one for unconstrained households. We assume that all households have high marginal propensities to consume with respect to changes in debt service payments (based on the discussion in Di Maggio et al. 2017) but a low marginal propensity to consume with respect to house and stock price appreciation (as discussed above).

Towards total consumption responses. By combining the ingredients in Table 1 we can construct our indirect estimates of cross-sectional consumption responses to monetary policy changes. We proceed as follows. For each household in our Survey of Consumer Finance dataset, we first classify them as financially constrained or unconstrained depending on their ratio of liquid assets to income, as already discussed above for Figure 1. Based on this classification we assign them the corresponding marginal propensities to consume reported in Table 1. Then, for each channel of policy transmission corresponding to a row in the table, we calculate their exposure to that channel. For example, a household's exposure to the stock market depends on the size of their stock holdings, while their exposure to mortgage rates depends on their current mortgage balance, and so on.[6] For labor income, we assume that low-income households are disproportionately exposed to the labor market, consistent with the findings in Guvenen et al. (2017). For each channel, we then compute a household's change in consumption as their exposure times the estimated price change (as listed in the table) times the assumed marginal propensity to consume. Summing across the different transmission channels, we arrive at a total effect on the household's consumption. We report the results in terms of a *percentage* change in consumption.[7]

Our calculation gives us an estimate of how each household in the Survey of Consumer Finances would change its consumption following a hypothetical monetary stimulus. We then summarize the distribution of these consumption changes in Figure 3. The figure shows the consumption effects across different quintiles of net worth (left panel), income (top right), and age (bottom right).

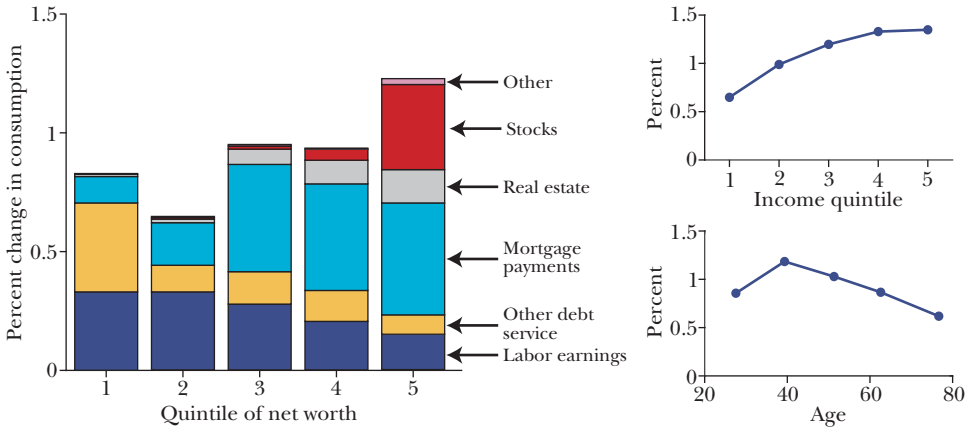
The left panel shows an average consumption change of 0.8 percent among low-wealth households and an average consumption change of 1.2 percent for wealthy households. The shaded areas in this left panel decompose the total consumption effect, revealing that labor income and nonmortgage debt drive the consumption response for low-wealth households, wages and mortgages are the main factors in the middle of the wealth distribution, and stock market gains are increasingly important at high wealth levels.

The top-right panel of Figure 3 instead shows the total consumption effect across income levels for working-age households. The average consumption responses within the top four quintiles are all between 1.0 and 1.3 percent, while

[6] To account for fixed-rate contracts and gradual refinancing we scale the change in mortgage rates and other interest rates by a factor of one half.

[7] To compute a percentage change, we need a baseline level of consumption. As the Survey of Consumer Finances does not report consumption, we impute it using the relationship between consumption and income in the Panel Study of Income Dynamics (Institute for Social Research 2019).

Figure 3
Indirect Calculation of Consumption Change across Quintiles of Net Worth, Income, and Age



Source: See online Appendix A for details on construction of sample and results.
Note: Sample restricted to households with net worth less than $2.5 million.

the lowest income quintile response is lower at 0.6 percent. Mechanically, a partial explanation for the insensitivity of consumption among the low-income group is that a substantial part (32 percent on average) of the income in this group is derived from social insurance and other transfers, which our analysis assumes is insensitive to monetary policy. Another potential explanation is that our analysis may actually understate the response of labor earnings in this group. Among households in the lowest income quintile, 34 percent had zero labor income during 2018, which was the reference year for the survey data we use. As our analysis "scales up" the existing income categories, the households that start with zero labor income will by construction not gain any labor income in our calculation; in practice, however, it may well be possible that these households would in fact enter employment in an expanding economy.

Finally, the bottom-right panel of Figure 3 shows the total consumption effect across age. Consistent with the logic discussed earlier (which suggested that young homeowners benefit substantially), we here find that the consumption effect peaks in early middle age and declines thereafter. Another factor driving the relatively small consumption response for old households is that a substantial part of their income comes from Social Security payments, which we assume are unaffected by monetary policy.

Overall, Figure 3 suggests two main takeaways on the cross-sectional incidence of monetary policy on household consumption. On the one hand, the incidence of the *individual channels* of monetary policy transmission to households is quite uneven. For example, mortgage payments and stocks have much stronger effects

at the top of the wealth distribution, while other debt services and labor income have stronger effects at the lower end. On the other hand, once aggregated across all channels, the *overall* consumption changes are much more evenly distributed. In particular, across the various cuts of the data, *all* groups materially increase their consumption. While there are some differences across groups, we view them overall as relatively modest. In particular, after a monetary stimulus that raises total GDP by 1 percent, even the least affected groups increase their consumption by a still material 0.6 percent (versus 1.3 percent for the most affected).

A more structural approach. While the indirect calculation above is attractive for its simplicity, it does require a large number of strong, reduced-form assumptions about household marginal propensities to consume. A more structural alternative is to use a model of household consumption decisions to infer how household consumption responds to the changes in income and prices induced by monetary policy. We do exactly that in McKay and Wolf (2022), using a general-equilibrium model in which households own a variety of long-duration assets and are unequally exposed to changes in labor earnings, in line with our earlier discussion of transmission channels. In that analysis, we also find that monetary stimulus leads to a quite evenly distributed increase in consumption across the population of households. Figure 4 summarizes our results, showing the consumption responses to monetary stimulus by net worth quintiles. The key takeaway is that, across all levels of wealth, consumption responds by nearly the same percentage amount. In this case, expansionary monetary policy roughly scales up everyone's consumption by the same amount as the aggregate, leaving each household's *share* of total consumption approximately unchanged.

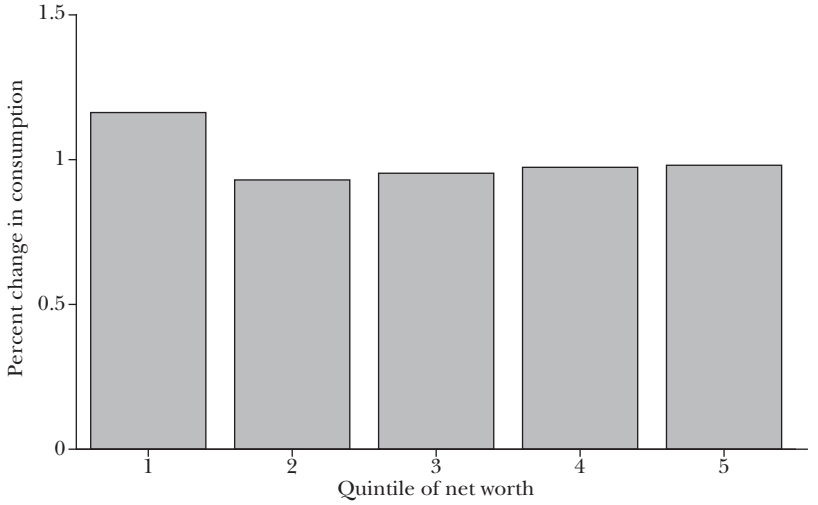
Inequality and the Aggregate Effects of Monetary Policy

Traditionally, the transmission of monetary policy to the macroeconomy has been analyzed in models populated by a representative household that chooses aggregate consumption, savings, and labor supply (for textbook treatments see Woodford 2003; Galí 2015). Recent research has pursued a different approach that explicitly incorporates household heterogeneity. This research agenda starts from microeconomic modeling of the choices of individual, heterogeneous households. We then arrive at predictions for aggregate variables by summing up across these heterogeneous households. This bottom-up, heterogeneous-agent approach can account for many of the distributional channels that we discussed in the previous section.

A new view on the channels of policy transmission. The heterogeneous-agent approach, with its emphasis on consumption-savings decisions at the household level, has changed our understanding of the *decomposition* of monetary policy effects into different underlying channels of transmission. Two margins of the transmission mechanism have received particular attention: the role of mortgage refinancing and the consumption response to changes in income.

Figure 4

Consumption Response to Monetary Stimulus at Different Levels of Net Worth from a General Equilibrium Model



Source: McKay and Wolf (2022).
Note: The figure shows results from a simulation of a monetary expansion that increases aggregate consumption by 1 percentage point.

As we have described above, expansionary monetary policy is typically associated with a decline in mortgage interest rates. Importantly, in the United States, homeowners often have the option to refinance their fixed-rate mortgages to take advantage of lower rates on new mortgages. As households refinance their mortgages to lower interest rates, their disposable income increases, allowing them to consume more. Propagation through such mortgage refinancing—rather than intertemporal substitution, as emphasized in traditional macroeconomic models—thus emerges as one of the most important direct transmission channels of monetary policy to consumer spending (Beraja et al. 2019).

In general equilibrium, the extra demand induced by monetary policy then translates to tighter labor markets, decreasing the unemployment rate and increasing labor incomes. Empirical evidence on the consumption response to changes in income shows that households spend quite strongly out of such temporary income gains (Johnson, Parker, and Souleles 2006). Many of the new heterogeneous-agent models of monetary policy transmission are designed to match these empirically estimated strong spending responses to changes in income. This leads to an important *indirect* channel of policy transmission: expansionary policy raises incomes and then households spend strongly out of that income, reinforcing the initial increase in demand. These indirect effects can be particularly strong due to a Keynesian multiplier logic: income increases spending, which then further increases household income, and so on (Kaplan, Moll, and Violante 2018; Auclert, Rognlie, and

Straub 2018; Bilbiie 2020; Patterson 2022). Important empirical support for these model predictions was provided in Holm, Paul, and Tischbirek (2021).

Reassessing the aggregate effects of monetary policy. The heterogeneous-agent view has changed our understanding of the precise channels through which monetary policy operates. Does this translate to a change in thinking about how monetary policy affects the macroeconomy as a whole? The answer is somewhat nuanced.

At a broad level, we would argue that our understanding of the effects of monetary policy on macro outcomes has not changed very much. Economists already have good *empirical* evidence on the average effects of monetary policy on macroeconomic outcomes (Ramey 2016). Any structural model of monetary policy transmission needs to be consistent with this evidence on the total effect, and so at best microeconomic heterogeneity will affect our understanding of the decomposition of this total effect into different channels. However, as we refine our understanding of the channels through which monetary policy operates, we may then also change our perspective on why policy may be more or less powerful at a given point in time depending on the state of the economy—a phenomenon known as "state dependence" in the effectiveness of policy. Such state dependence is difficult to identify from purely empirical analysis of time series data because it requires the researcher to estimate the effects of policy in different states of the economy (as opposed to just estimating some average effect). Structural modeling—and in particular modeling that carefully accounts for microeconomic channels of transmission—is thus the most promising avenue to learn about such state dependence.

The role of mortgage refinancing. One likely reason for state dependence in the aggregate effects of monetary policy is related to household mortgage refinancing (Berger et al. 2021; Eichenbaum, Rebelo, and Wong 2022). The incentives for households to refinance their mortgages depend on the difference between the mortgage rate offered on a new loan and their existing interest rates. If households currently have high interest rates on their mortgages, then they will be likely to refinance soon anyway, and thus any additional changes in mortgage rates due to monetary policy will flow through strongly to the rates households actually pay. On the other hand, if households are already paying low interest rates, they will be less likely to refinance, and so any marginal rate change related to monetary policy will have a smaller impact on the rates households are actually paying.

Beraja et al. (2019) analyze another related reason for why the strength of the mortgage refinancing channel is likely to vary over time. In order to refinance a mortgage, the homeowner must be approved for a new loan. Importantly, obtaining a new loan can be more or less difficult depending on a number of factors that are likely to vary over time. For example, some lenders require that the new loan is for no more than 80 percent of the home's value. In the aftermath of the Great Recession, declines in home prices left many homeowners unable to meet this requirement. These homeowners found it difficult to refinance their mortgages and were unable to take advantage of the low interest rates offered on new mortgages. In the aggregate, at times when many households are in this situation, the mortgage channel of monetary policy transmission will be muted.

Spending on durables. Time variation in household demand for durable goods is another reason for why the sensitivity of the economy to monetary policy may change over time (Berger and Vavra 2015; Tenreyro and Thwaites 2016). To make this more concrete, consider a household that is contemplating the purchase of a new car. For such a household, a change in interest rates could make the difference between buying the car and not buying the car. It follows that monetary policy is likely to have large effects when many households are contemplating such purchases, as usually happens in times of economic expansion. Conversely, in a downturn, few households are contemplating any big purchases, and so monetary policy transmission may be weakened.

Intertemporal shifting of demand. Finally, in addition to raising the possibility that the power of monetary policy varies over time, the recent heterogeneous-household research agenda has also raised questions about the medium-term effects of policy. We typically think that monetary stimulus raises demand in the short run. But what about at longer horizons? Mian, Straub, and Sufi (2021) as well as McKay and Wieland (2021) highlight forces whereby monetary stimulus raises demand in the near term but depresses it at longer horizons. The logic of these arguments is that monetary stimulus raises demand today, but changes household balance sheets in ways that leave them less willing to spend in the future. This outcome could occur because households take on additional debt (Mian, Straub, and Sufi 2021) or because they purchase durable goods (McKay and Wieland 2021). These studies predict that changes in interest rates will tend to be persistent, because stimulus today requires continued stimulus in the future to offset the endogenous reduction in future demand.

Taking stock. Our overall conclusion is that recent research emphasizing microeconomic household heterogeneity has led to an evolution—rather than a revolution—in our understanding of the aggregate effects of monetary policy. Compared to prior work, this research places emphasis on a different set of channels shaping the aggregate effects of policy changes. While these channels introduce some novel sources of state dependence and intertemporal shifting of demand, the existing empirical evidence on the aggregate short-run effects of monetary policy remains an important touchstone for both representative-agent as well as heterogeneous-agent models.

Optimal Monetary Policy with Household Inequality

So far, we have discussed the interaction between monetary policy and inequality on purely positive grounds, asking whether (1) monetary policy affects the evolution of inequality and (2) inequality affects the propagation of monetary policy. We now turn to a normative question, asking how inequality may change our view of *optimal* monetary policy.

In the long run, real economic outcomes—including inequality across households—are largely outside the control of monetary policy. For optimal monetary

policy, inflation stabilization remains the only long-run consideration. Here, we will focus instead on how monetary policy should respond to fluctuations in the economy in the short run. We will start by considering a central bank with a narrow mandate, seeking only to stabilize macroeconomic aggregates. We then consider how a central bank with a broader mandate that includes distributional concerns would act differently. Much of the intuitive discussion in this section is based on our formal analysis in McKay and Wolf (2022).

A Narrow Mandate

Traditionally, central banks have pursued the dual objectives of stabilizing (1) inflation as well as (2) real aggregate activity measures (such as employment or GDP). Will household inequality affect the behavior of a central bank with these targets?

Policy problem. Figure 5 illustrates the policy problem faced by a central bank with a mandate to stabilize inflation and aggregate output. The top panel begins by showing policymaker preferences (in orange) and constraints (in blue). The figure features output on the horizontal axis and inflation on the vertical axis, with the output-inflation pair marked as (y^*, π^*) as the policymaker's desired outcome. The orange curve then shows an indifference curve corresponding to policymaker preferences, with better outcomes closer to the target.[8] The blue line is the constraint set—the set of inflation-output pairs that the policymaker can in principle implement. The line corresponds to an aggregate supply curve, with its upward slope reflecting the usual logic that higher utilization of the economy's resources leads to upward pressure on costs and thus prices. Intuitively, if the economy is going to produce more output, then workers will have to be incentivized to work longer hours and wages and prices will increase. The bottom panel of the figure on the other hand represents the aggregate demand block of the economy: to achieve a given amount of real activity (output, again on the horizontal axis), real interest rates (nominal rates less inflation, shown on the vertical axis) need to be set at a certain level. The line is downward-sloping, reflecting the idea that higher interest rates depress aggregate demand, as discussed earlier.

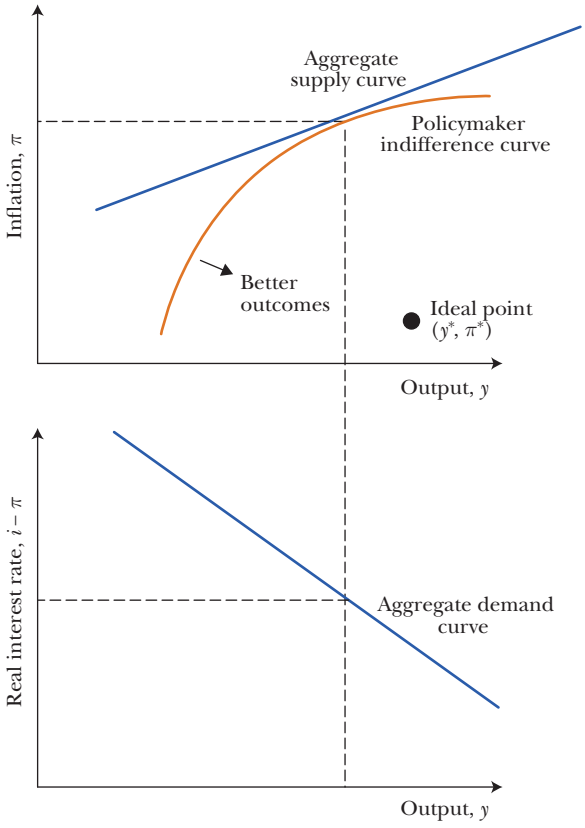
The policy problem is to choose the nominal interest rate so as to minimize the deviations of output and inflation from their target values. As the policymaker changes the nominal interest rate, the real interest rate changes too, moving the economy along the aggregate demand curve (bottom panel) to determine the level of output. Moving to the upper panel, the aggregate supply curve determines the associated level of inflation. The solution to this optimal policy problem is

[8] Central bank preferences are often described by a loss function that captures the idea that the central bank dislikes it when economic outcomes differ from the targeted outcomes. For example, the indifference curve we are plotting corresponds to the loss function

$$(y - y^*)^2 + (\pi - \pi^*)^2,$$

where y is output, π is inflation, and the starred variables are the targets for output and inflation. Unlike consumer theory where we maximize utility, here we want to minimize the loss function.

Figure 5
Optimal Monetary Policy with a Narrow Mandate



straightforward: the monetary policymaker focuses on the upper panel of the figure and simply chooses the best feasible output-inflation pair. Visually, the optimal pair is given by the point of tangency with the indifference curve. The policymaker then uses the aggregate demand curve to determine which nominal interest rate to set in order to arrive at the desired level of output. The solution is shown by the dashed lines.

The role of inequality. How might household inequality affect this policy problem? Remember that we are assuming (for the moment) that inequality is not a target of the central bank, which implies that the policymaker indifference curve is not affected by inequality. Through the lens of the simple framework shown in Figure 5, most of the research on the connection between monetary policy and inequality discussed in the previous sections may be interpreted as studying ways in which household inequality could change the economy's aggregate demand relationship—that is, the mapping from interest rates to aggregate demand shown in the lower panel. However, the framing of the problem in the figure shows that

changes in the demand block alone will not affect the optimal output and inflation outcomes, simply because the optimal policy choice is already fully pinned down by policymaker preferences and the supply side of the economy. It follows that changes in aggregate demand due to inequality will not affect the optimal inflation and output levels, though they may affect the nominal interest rate required to implement this optimal output-inflation allocation.

How big are those effects on optimal interest rates likely to be? Graphically, for a given output-inflation outcome, the optimal interest rate is determined by the *slope* and *intercept* of the economy's aggregate demand relationship. The slope of the curve reflects the sensitivity of aggregate demand to changes in real interest rates. As we discussed in the previous section, the heterogeneous-agent view of monetary policy transmission has not materially changed our broad understanding of this sensitivity. The intercept of the line, on the other hand, reflects forces that determine aggregate demand at any given interest rate, with changes in this intercept reflecting so-called "aggregate demand shocks." Many plausible examples of such shocks are explicitly distributional in nature. For example, a tightening of credit conditions could require borrower households to reduce their debt levels, thus leading to a reduction in their spending and thus aggregate demand (for example, Guerrieri and Lorenzoni 2017). Alternatively, fiscal transfer payments to financially constrained households could lead to an increase in total consumer spending (for example, Wolf 2021). The theory sketched here suggests that such shifts in aggregate demand would lead to an equilibrium adjustment in interest rates while leaving optimal inflation and output outcomes unchanged.

A Broad Distributional Mandate for Monetary Policy

We now consider a central bank that explicitly incorporates distributional concerns as one of its policy goals, presumably along with its traditional output and inflation goals.[9] Long-run trends in inequality of course primarily reflect economic forces unrelated to monetary policy and the business cycle at large. However, short-run business-cycle fluctuations may well have material (short-term) effects on inequality, simply because aggregate shocks need not affect everyone in the same way. If a central bank's mandate includes distributional outcomes, then it will try to set its policy in a way that redistributes towards the hardest-hit households, essentially providing some insurance to those most exposed to aggregate shocks.

The role of insurance. In an ideal world, households would be able to buy insurance against all types of adverse events—including aggregate cyclical fluctuations—in private markets. For example, a worker could buy an insurance policy against the risk of unemployment. With such perfect insurance markets, standard macroeconomic models would predict that the consumption of all households

[9] Studies of optimal monetary policy that incorporate distributional effects include Bhandari et al. (2021), Acharya, Challe (2020), Le Grand, Martin-Baillon, and Ragot (2021), Dávila and Schaab (2022), and McKay and Wolf (2022). In the latter study, we develop the views we describe in this section more formally.

would move up and down in proportion to aggregate consumption. Due to issues of moral hazard and adverse selection, however, many such insurance markets do not exist. With imperfect insurance markets, the economy will move away from this efficient pattern of risk sharing. Thus, following an aggregate shock, some households may be more severely affected than others and therefore reduce their consumption by more than the rest.

The social insurance benefit of filling in for these missing markets is widely recognized in other areas of public policy, including discussions of unemployment insurance systems, tax policy, and social safety net programs. Similar underlying concerns may thus also guide optimal monetary policymaking: interest rates may be set in a way to both move aggregate consumption to the desired overall level *and* to smooth out consumption changes across households, essentially moving the cross-sectional consumption distribution closer to the desired efficient pattern of risk sharing.

Can monetary policy provide insurance? The evidence that we reviewed above is informative about the extent to which monetary policy can provide insurance and thus achieve such distributional objectives. Our main conclusion from that discussion was that monetary policy has rather evenly distributed effects across different groups of households—that is, expansionary monetary policy scales up the consumption of different households by similar proportions. To make the argument particularly stark, suppose for a moment that monetary policy was exactly distributionally neutral in the sense of scaling everyone's consumption up and down in perfect unison. In that case, monetary policy interventions would not bring us any closer to the efficient risk-sharing outcome, and so social insurance would not be a consideration for optimal monetary policy. Intuitively, even if a monetary policymaker would like to lean against inequality, monetary policy is not well-suited to do so, and so the monetary policymaker will act *as if* it had only a narrow mandate.

The actual situation is of course not as extreme as this—the effects of policy on consumption are not exactly equal in percentage terms—so there is some scope for monetary policy to alter the distribution of consumption. However, given the modest extent of these distributional effects, large changes in monetary policy would be needed to have a substantial effect on the consumption distribution. Such large changes would likely be costly in terms of other policy goals (notably aggregate output and inflation stabilization). As a result, a central bank that targets both conventional aggregate outcomes as well as distributional outcomes is unlikely to deviate too much from the policies pursued by a central bank with a narrow mandate that just targets aggregate outcomes.

Conclusion

In this paper, we have taken stock of the recent research agenda that studies the connections between monetary policy and inequality, with three main conclusions. First, our reading of the empirical evidence suggests that monetary policy has

a relatively uniform incidence across households. Second, accounting for micro heterogeneity across households changes our understanding of the transmission channels for monetary policy. It has not, however, changed our understanding of the broad patterns of how monetary policy effects the macroeconomy. Third, our first two observations taken together somewhat limit the scope to which household inequality is likely to affect optimal monetary policy design, even if the central bank has a broad mandate that includes distributional considerations.

However, we emphasize that these broad conclusions come with important qualifiers, which we view as topics for future research. First, in keeping with the recent academic literature on inequality and monetary policy, our analysis throughout this article focused on how heterogeneity interacts with the demand side of the economy. Changes to the supply side would affect optimal outcomes even for a central bank with a conventional narrow mandate. Second, our discussion has omitted the heterogeneity in consumption baskets across households. There is, however, evidence that low-income groups and racial minorities consume goods with more volatile prices (for example, Cravino, Lan, and Levchenko 2020; Lee, Macaluso, and Schwartzman 2021), thus adding a further possible layer to the distributional effects of monetary policy. Finally, our conclusions on the distributional effects of monetary policy remain tentative, relying either on noisy consumption measures or assumptions on household consumption-savings decisions. More empirical work on these topics would be very welcome.

■ *The authors thank Adrien Auclert, James Cloyne, Martin Blomhoff Holm, Pascal Paul, Joe Vavra, Johannes Wieland, and Abbie Wozniak for helpful comments.*

References

Acharya, Sushant, Edouard Challe, and Keshav Dogra. 2020. "Optimal Monetary Policy According to HANK." CEPR Discussion Paper 14429.

Amberg, Niklas, Thomas Jansson, Mathias Klein, and Anna Rogantini Picco. 2021. "Five Facts about the Distributional Income Effects of Monetary Policy." CESifo Working Paper.

Andersen, Asger Lau, Niels Johannesen, Mia Jørgensen, and José-Luis Peydro. 2021. "Monetary Policy and Inequality." Universitat Pompeu Fabra Working Paper 1761.

Auclert, Adrien. 2019. "Monetary Policy and the Redistribution Channel." *American Economic Review* 109 (6): 2333–67.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2018. "The Intertemporal Keynesian Cross." NBER Working Paper 25020.

Bauer, Michael D., and Eric T. Swanson. 2022. "A Reassessment of Monetary Policy Surprises and High-Frequency Identification." NBER Working Paper 29939.

Beraja, Martin, Andreas Fuster, Andreas Hurst, and Joseph Vavra. 2019. "Regional Heterogeneity and the Refinancing Channel of Monetary Policy." *Quarterly Journal of Economics* 134 (1): 109–83.

Berger, David, Veronica Guerrieri, Guido Lorenzoni, and Joseph Vavra. 2018. "House Prices and Consumer Spending." *Review of Economic Studies*, 85 (3): 1502–42.

Berger, David, Konstantin Milbradt, Fabrice Tourre, and Joseph Vavra. 2021. "Mortgage Prepayment and Path-Dependent Effects of Monetary Policy." *American Economic Review* 111 (9): 2829–78.

Berger, David, and Joseph Vavra. 2015. "Consumption Dynamics during Recessions." *Econometrica* 83 (1): 101–54.

Bernanke, Ben S., and Kenneth N. Kuttner. 2005. "What Explains the Stock Market's Reaction to Federal Reserve Policy?" *Journal of Finance* 60 (3): 1221–57.

Best, Michael C., James S. Cloyne, Ethan Ilzetzki, and Henrik J. Kleven. 2020. "Estimating the Elasticity of Intertemporal Substitution Using Mortgage Notches." *Review of Economic Studies* 87 (2): 656–90.

Bhandari, Anmol, David Evans, Mikhail Golosov, and Thomas J. Sargent. 2021. "Inequality, Business Cycles, and Monetary-Fiscal Policy." *Econometrica* 89 (6): 2559–99.

Bilbiie, Florin O. 2020. "The New Keynesian Cross." *Journal of Monetary Economics* 114: 90–108.

Chang, Minsu, and Frank Schorfheide. 2022. "On the Effects of Monetary Policy Shocks on Earnings and Consumption Heterogeneity." CEPR Discussion Paper 17049.

Chodorow-Reich, Gabriel, Plamen T. Nenov, and Alp Simsek. 2021. "Stock Market Wealth and the Real Economy: A Local Labor Market Approach." *American Economic Review* 111 (5): 1613–57.

Cloyne, James, Clodomiro Ferreira, and Paolo Surico. 2020. "Monetary Policy When Households Have Debt: New Evidence on the Transmission Mechanism." *Review of Economic Studies* 87 (1): 102–29.

Coibion, Olivier, Yuriy Gorodnichenko, Lorenz Kueng, and John Silvia. 2017. "Innocent Bystanders? Monetary Policy and Inequality." *Journal of Monetary Economics* 88: 70–89.

Cravino, Javier, Ting Lan, and Andrei A. Levchenko. 2020. "Price Stickiness along the Income Distribution and the Effects of Monetary Policy." *Journal of Monetary Economics* 110: 19–32.

Dávila, Eduardo, and Andreas Schaab. 2022. "Optimal Monetary Policy with Heterogeneous Agents: A Timeless Ramsey Approach." https://www.eduardodavila.com/research/davila_schaab_timeless_ramsey.pdf.

Di Maggio, Marco, Amir Kermani, Benjamin J. Keys, Tomasz Piskorski, Rodney Ramcharan, Ami Seru, and Vincent Yao. 2017. "Interest Rate Pass-Through: Mortgage Rates, Household Consumption, and Voluntary Deleveraging." *American Economic Review* 107 (11): 3550–88.

Doepke, Matthias, and Martin Schneider. 2006. "Inflation and the Redistribution of Nominal Wealth." *Journal of Political Economy* 114 (6): 1069–97.

Eichenbaum, Martin, Sergio Rebelo, and Arlene Wong. 2022. "State-Dependent Effects of Monetary Policy: The Refinancing Channel." *American Economic Review* 112 (3): 721–61.

Fagereng, Andreas, Matthieu Gomez, Émilien Gouin-Bonenfant, Martin Holm, Benjamin Moll, and Gisle Natvik. 2022. "Asset-Price Redistribution." https://www.matthieugomez.com/files/apr.pdf.

Galí, Jordi. 2015. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*. Princeton, NJ: Princeton University Press.

Gertler, Mark, and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7 (1): 44–76.

Greenwald, Daniel L., Matteo Leombroni, Hanno Lustig, and Stijn Van Nieuwerburgh. 2021. "Financial and Total Wealth Inequality with Declining Interest Rates." NBER Working Paper 28613.

Guerrieri, Veronica, and Guido Lorenzoni. 2017. "Credit Crises, Precautionary Savings, and the Liquidity Trap." *Quarterly Journal of Economics* 132 (3): 1427–67.

Guren, Adam M., Alisdair McKay, Emi Nakamura, and Jon Steinsson. 2021. "Housing Wealth Effects: The Long View." *Review of Economic Studies* 88 (2): 669–707.

Guvenen, Fatih, Serdar Ozkan, and Jae Song. 2014. "The Nature of Countercyclical Income Risk." *Journal of Political Economy* 122 (3): 621–60.

Guvenen, Fatih, Sam Schulhofer-Wohl, Jae Song, and Motohiro Yogo. 2017. "Worker Betas: Five Facts about Systematic Earnings Risk." *American Economic Review* 107 (5): 398–403.

Holm, Martin B., Pascal Paul, and Andreas Tischbirek. 2021. "The Transmission of Monetary Policy under the Microscope." *Journal of Political Economy* 129 (10): 2861–904.

Iacoviello, Matteo. 2005. "House Prices, Borrowing Constraints, and Monetary Policy in the Business Cycle." *American Economic Review* 95 (3): 739–64.

Johnson, David S., Jonathan A. Parker, and Nicholas S. Souleles. 2006. "Household Expenditure and the Income Tax Rebates of 2001." *American Economic Review* 96 (5): 1589–610.

Kaplan, Greg, Giovanni L. Violante, and Justin Weidner. 2014. "The Wealthy Hand-to-Mouth." *Brookings Papers on Economic Activity* 45 (1): 77–138.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. "Monetary Policy According to HANK." *American Economic Review* 108 (3): 697–743.

Le Grand, François, Alaïs Martin-Baillon, and Xavier Ragot. 2021. "Should Monetary Policy Care about Redistribution? Optimal Fiscal and Monetary Policy with Heterogeneous Agents." http://xavier-ragot.fr/pdf/progress/LeGrand_Ragot_Money.pdf.

Lee, Munseob, Claudia Macaluso, and Felipe Schwartzman. 2021. "Minority Unemployment, Inflation, and Monetary Policy." https://www.minneapolisfed.org/~/media/assets/events/2021/fall-2021-institute-research-conference/claudia-macaluso_minority-unemployment-and-monetary-policy.pdf?la=en.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H. Stock. 2014. "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve." *Journal of Economic Literature* 52 (1): 124–88.

McKay, Alisdair, and Johannes F. Wieland. 2021. "Lumpy Durable Consumption Demand and the Limited Ammunition of Monetary Policy." *Econometrica* 89 (6): 2717–49.

McKay, Alisdair, and Christian K. Wolf. 2022. "Optimal Policy Rules in HANK." https://alisdairmckay.com/pdfs/hank_optpol.pdf.

McKay, Alisdair, and Christian K. Wolf. 2023. "Replication data for: Monetary Policy and Inequality." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E183663V1.

Mian, Atif, Kamalesh Rao, and Amir Sufi. 2013. "Household Balance Sheets, Consumption, and the Economic Slump." *Quarterly Journal of Economics* 128 (4): 1687–726.

Mian, Atif, Ludwig Straub, and Amir Sufi. 2021. "Indebted Demand." *Quarterly Journal of Economics* 136 (4): 2243–307.

Okun, Arthur M. 1973. "Upward Mobility in a High-Pressure Economy." *Brookings Papers on Economic Activity* 4 (1): 207–62.

Patterson, Christina. 2022. "The Matching Multiplier and the Amplification of Recessions." https://www2.census.gov/ces/wp/2022/CES-WP-22-20.pdf.

Plagborg-Møller, Mikkel, and Christian K. Wolf. 2021. "Local Projections and VARs Estimate the Same Impulse Responses." *Econometrica* 89 (2): 955–80.

Ramey, V.A. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics*, Vol. 2, edited by John B. Taylor and Harald Uhlig, 71–162. Amsterdam: North-Holland.

Slacalek, Jiri, Oreste Tristani, and Giovanni L. Violante. 2020. "Household Balance Sheet Channels of Monetary Policy: A Back of the Envelope Calculation for the Euro Area." *Journal of Economic Dynamics and Control* 115: 103879.

Tenreyro, Silvana and Greg Thwaites. 2016. "Pushing on a String: US Monetary Policy Is Less Powerful in Recessions." *American Economic Journal: Macroeconomics* 8 (4): 43–74.

Tracy, Joseph, and Joshua Wright. 2012. "Why Mortgage Refinancing Is Not a Zero-Sum Game." *Liberty Street Economics*, January 11. https://libertystreeteconomics.newyorkfed.org/2012/01/why-mortgage-refinancing-is-not-a-zero-sum-game/.

Wolf, Christian K. 2021. "Interest Rate Cuts vs. Stimulus Payments: An Equivalence Result." NBER Working Paper 29193.

Wong, Arlene. 2021. "Refinancing and the Transmission of Monetary Policy to Consumption." https://static1.squarespace.com/static/576576adbe659449f97e0d35/t/60770bfec442b6095624f676/1618414592249/Paper_AERrevision2.pdf.

Woodford, Michael. 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton, NJ: Princeton University Press.

Journal of Economic Perspectives—Volume 37, Number 1—Winter 2023—Pages 145–168

Unraveling the Hispanic Health Paradox

José Fernandez, Mónica García-Pérez, and Sandra Orozco-Aleman

Hispanics in the United States tend to have lower household income, education, and health insurance coverage when compared to non-Hispanic Whites. Despite these economic disadvantages, paradoxically, Hispanics have displayed an equality with or even advantages over other minority groups and non-Hispanic Whites across a wide range of health outcomes. For example, in 2019, the Hispanic population had a life expectancy advantage of 3.0 years over the non-Hispanic White population and 7.1 years relative to the non-Hispanic Black population, despite having real household income that was 26 percent lower than non-Hispanic White households (Wilson 2020). Hispanic immigrants have also shown lower infant mortality rates and prevalence of mental illnesses. These stylized facts are collectively known as the "Hispanic health paradox." This essay will provide an overview of the Hispanic health paradox literature. We will document different instances of the Hispanic health paradox across various measures: life expectancy at birth, infant mortality rate, death rates, causes of death, and morbidity. We will discuss the leading explanations of the Hispanic health paradox and possible ways for economists to contribute to this discussion.

The origin of the Hispanic health paradox is often traced to the seminal paper by Markides and Hazuda (1980), in which the outperformance of Hispanics was

■ *José Fernandez is Associate Professor of Economics, University of Louisville, Louisville, Kentucky. Mónica García-Pérez is Professor of Economics, St. Cloud State University, St. Cloud, Minnesota. Sandra Orozco-Aleman is Associate Professor of Economics, Mississippi State University, Mississippi State, Mississippi. Their email addresses are jose.fernandez@louisville.edu, migarciaperez@stcloudstate.edu, and sorozco@business.msstate.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.1.145.

deemed an "epidemiological paradox." The authors found that Mexican Americans in southwest Texas had a lower infant mortality rate relative to other groups, including non-Hispanic Whites. Markides and Coreil (1986) reported the same phenomenon for life expectancy, mortality, disease-related health outcomes, and mental and functional health. In fact, the first Hispanic health advantage reported was observed for mental health among Mexican Americans (Karno and Edgerton 1969). Markides and Eschbach (2005) renamed these advantages the Hispanic paradox. In their discussion, they highlight the role of immigration in explaining the paradox, with the initial assumption that immigrants need to be healthy enough to endure the cost associated with immigration: travel, adaptation to new customs, new laws, and potentially with few resources or support available. The Hispanic health paradox is closely tied to the "healthy immigrant effect" (also known as the "healthy immigrant paradox"). The healthy immigrant effect is an observed time path in which the health of immigrants just after the migration is substantially better than that of comparable native-born people, but worsens with additional years in the new country. Stephen et al. (1994) were the first to identify this effect using the 1989 National Health Interview Survey, because this was the first time the survey incorporated the number of years lived in the country. Since then, various authors have identified this effect across numerous health outcomes (Ali 2002; Goel et al. 2004; Kennedy et al. 2015; McDonald and Kennedy 2004; Puyat 2013; Wu and Schimmele 2005; Jasso et al. 2005; Roger et al. 2011; Constant et al. 2018).

A vast majority of the Hispanic health paradox literature has treated Hispanics in the United States as a monolithic group.[1] Leading explanations of the Hispanic health paradox can be different due to backgrounds and characteristics. To that end, whenever possible, we disaggregate our findings by nativity and ancestry.

The Paradox in the Health Statistics

The Hispanic health paradox manifests itself through a variety of different health criteria: life expectancy, death rates, infant mortality, leading causes of death, and morbidity. By looking at differences across these measures, how the measures have been evolving, and differences across Hispanic subgroups (where such information is available), we can begin to explore some possible reasons behind the paradox itself.

The data sources selected are based on three criteria. First, we use nationally representative sources widely used in the literature discussing the paradox. Second,

[1] We use Hispanics throughout the document to be consistent with the word used in most US government surveys, but we recognize there are differences between the groups identified as Latinos (or sometimes Latinx), which refers to the country of origin in Latin America, versus Hispanics, which refers to a Spanish-speaking country of origin.

sample sizes are large enough to allow us to disaggregate Hispanics into subgroups by ancestry and nativity. Third, the data we use are publicly accessible, so those interested in this literature can pursue their research interests.

Life Expectancy at Birth

The Hispanic health paradox is perhaps most prevalent when discussing life expectancy and infant mortality (Markides and Eschbach 2011). Since 2006, Hispanics have had the highest life expectancy at birth of all groups despite their disadvantaged socioeconomic profile.[2] Life expectancy at birth was 81.8 years for the Hispanic population in 2019, 78.8 for non-Hispanic whites, and 74.9 for the non-Hispanic Black. Between 2006 and 2019, life expectancy increased by 1.2 years for Hispanics, 0.7 for the non-Hispanic White population, and 2.0 for the non-Hispanic Black population. To put these gains in perspective, life expectancy at birth in the United States increased by more than eleven years between 1960 and 2019, going from 69.7 to 81.8 years.[3]

Life expectancy is defined as the average number of years of life remaining for a person at a particular age. Data used to calculate life expectancy include death counts and US Census population estimates.[4] Death counts are obtained from death certificates reported to the National Center for Health Statistics (NCHS) as part of the National Vital Statistics System (NVSS). Death certificates include information on the race and Hispanic origin of the deceased. Funeral directors collect information about ethnicity from family members of the deceased or from hospital records. While life expectancy is an important indicator of the population's health, the analysis of the Hispanic population must take into account that life expectancy tables do not consider the possibility that some Hispanic deaths are not accounted for due to return migration.

Death Rates

The unadjusted death rate is the total number of deaths per 100,000 population. The unadjusted rates are sensitive to differences in age profiles across populations. Because mortality rates increase with age, a higher mortality rate might simply reflect that the population is older. Mortality rates can be standardized using a weighted average of the age-specific mortality rates to eliminate the effect of different age distributions among different populations. The age-adjusted death rates should be

[2]The first year for comparison is 2006. Estimates calculated before that year are considered unreliable due to quality issues associated with race and Hispanic origin misclassification on US death certificates, leading to underestimating death rates for Hispanics. Additionally, a misstatement of age in vital statistics and census data at the oldest ages observed before 2006 led to underestimating mortality at the oldest ages (Arias 2010).

[3]The US Census Bureau produced population tables in which data for multiple-race persons were bridged back to single-race categories. Life expectancy at birth in 2006 and 2019 is shown in Table A1 in the Appendix.

[4]Population data used to calculate life expectancy in 2006 and 2019 were based on the 2000 and 2010 census counts, respectively. The life expectancy calculation in 2006 also used Medicare data as it was considered more reliable for estimating mortality at the oldest ages as it requires proof of age.

viewed as relative indexes rather than actual measures of mortality risk because they compare the risk of death among two populations with the counterfactual assumption that both groups have the same age distribution. Data used to calculate death rates comes from death certificates and US Census population estimates.

As shown in the left-hand panel of Figure 1, Hispanics have lower unadjusted death rates than the non-Hispanic White and non-Hispanic Black populations. In 2019, the unadjusted death rates were 3.1 and 2.3 times greater for non-Hispanic Whites (1,090 per 100,000) and non-Hispanic Blacks (807) than for Hispanics (351). Across Hispanic subgroups, Americans of Cuban origin have the highest unadjusted mortality rate at 716 deaths per 100,000, followed by Puerto Ricans (466), Mexicans (302), Central Americans (205), and South Americans (246).[5]

The right-hand panel of Figure 1 shows Hispanics also have lower age-adjusted death rates than the non-Hispanic White and non-Hispanic Black populations, although the gaps are not as dramatic. In 2019, age-adjusted death rates were 1.7 and 1.4 times greater for the non-Hispanic Black (871) and non-Hispanic White (737 per 100,000) populations than for the Hispanic population (524). In particular, notice that the age-adjusted rate for non-Hispanic White and Cubans dropped significantly, indicating potential differences in the age distribution of those groups. Moreover, those results are consistent for males and females.[6]

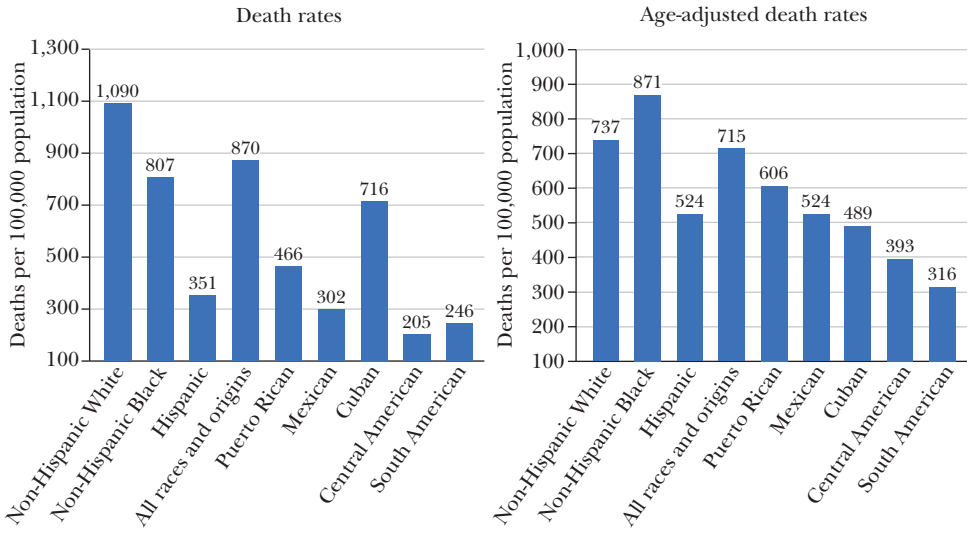
Among Hispanic subgroups (the members of which are self-reported and based on ancestry), Puerto Ricans have the highest mortality rate at 605.7 deaths per 100,000, followed by Mexicans (523.7), Cubans (489.1), Central Americans (393.2), and South Americans (315.5). The age-adjusted rate for Cubans is now below the average rate for Hispanics—a dramatic change, as Cubans had the highest unadjusted death rate of all Hispanic subgroups.

The significant differences in adjusted and unadjusted rates highlight the importance of analyzing age distribution differences among Hispanic subgroups; in turn, these differences can help to illuminate the mechanisms that can contribute to the existence of the Hispanic health paradox. While Mexicans, Central Americans, and Puerto Ricans have a higher proportion of individuals aged 45 or below (75, 75, and 70 percent, respectively), the Cuban and non-Hispanic White populations have a higher proportion of individuals aged 45 and above (45 and 49 percent, respectively). Similarly, we find significant differences across Hispanic subgroups for the average age at death. In 2019, the Cuban population had the highest average age at death, with 77.6 years, followed by the non-Hispanic White with 75.1 years. On the other hand, the groups with the lowest average age at death include the non-Hispanic Black at 65.9 years, Mexicans at 64.2, and Central Americans at 60.3 years. As we will discuss later in the paper, these differences are affected by age-selective migration, and perhaps especially by differences across Hispanic subgroups in the

[5]Estimates show that mortality among Hispanics may be understated due to the net underreporting of Hispanic origin on the death certificate by approximately 3 percent. However, misclassification of Hispanic origin on the death certificate is relatively stable across age groups (Xu et al. 2021).

[6]Age-adjusted death rates are further disaggregated by gender in the Appendix, Figure A1.

Figure 1
Unadjusted and Adjusted Death Rates



Source: Xu et al. (2021).
Notes: Death rates are deaths per 100,000 population. Mortality data is from the National Vital Statistics System (death certificates) and US Census population estimates. Estimates for males and females are shown in Figure A1 in the online Appendix.

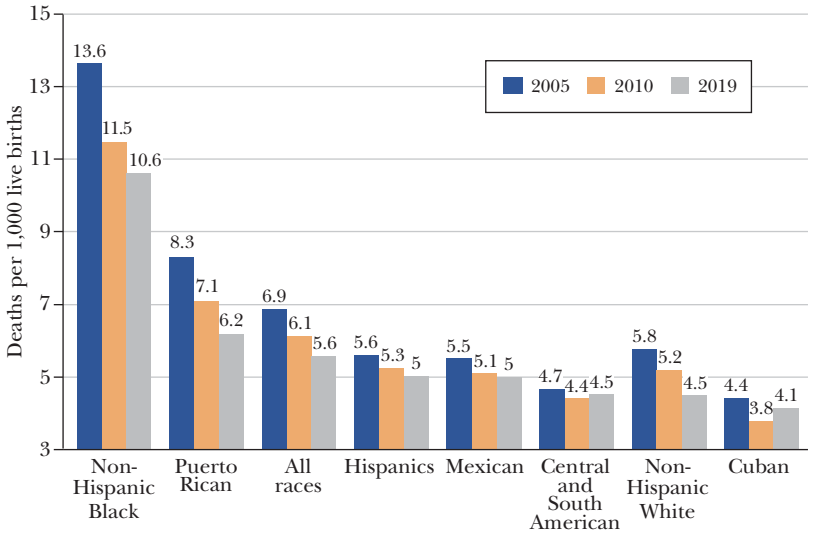
likelihood of those who are older or in poor health to return to their home country. Specifically, Mexican and Central American populations are more likely to return to their home country than immigrants from countries in South America and Cuba (Arenas et al. 2015).

While accounting for age differences across Hispanic subgroups helps explain the sizable raw death rate differentials, health within an age group can still be affected. The return migration of less healthy immigrants to Mexico and Central America (relative to South America and Cuba) would imply that older Hispanics are healthier than non-Hispanic Whites of the same age.

Infant Mortality Rate

Infant mortality rates are calculated as the number of deaths per 1,000 live births (aged <1 year) in the specified group. Data used to calculate infant mortality rates comes from the National Center for Health Statistics (NCHS) linked birth/infant death files and not from birth certificates. As part of the Vital Statistics Cooperative Program (VSCP), each state links information from the birth and death certificate for each infant (aged <1 year) who dies in the United States. The linked birth/infant death data include individuals born in the 50 states and Washington, DC, and maternal ethnicity and nationality are self-reported. For Hispanics, the

Figure 2
Infant Mortality Rate



Sources: Ely and Driscoll (2021), Mathews and MacDorman (2013), and MacDorman and Mathews (2013).
Note: Infant mortality rates are calculated as the number of infant deaths per 1,000 live births in the specified group. Data come from the NCHS linked birth/infant death datasets.

data distinguish six Hispanic groups by place of origin: Mexico, Puerto Rico, Cuba, Central America, South America, and other or unknown origins. Data only include the deaths of infants who were born and died in the United States. These data miss foreign-born deaths, although those deaths appear in the raw mortality files.

The literature on infant mortality rate has found favorable infant survival rates for some Asian and Hispanic groups attributable to a high percentage of births to immigrant women—women characterized as having lower infant mortality than native-born mothers—as well as to sociodemographic, behavioral, maternal health, and birth outcome risk factors. For example, lower mortality rates of Central and South American mothers have been attributed to the large concentration of births to foreign-born women from those groups. Likewise, foreign-born Mexican American women exhibit less risky health profiles than US-born Mexican American women, explaining their lower infant mortality rates (Hummer et al. 1999).

Figure 2 shows that the mortality rate in 2019 was 5.6 infant deaths per 1,000 live births, a historic low for the country. The infant mortality rate for infants of Hispanic women (5.0) is less than half the rate for non-Hispanic Black women (10.6), women who might have similar socioeconomic conditions, and only slightly above the mortality rate for infants of non-Hispanic White women (4.5).

Data can be divided further into Hispanic-origin subgroups: specifically, Mexican, Puerto Rican, Central and South American, and Cuban, in addition to the

residual category of other Hispanics. As shown in Figure 2, infants born to Puerto Rican women had the highest mortality rate (6.2 per 1,000 live births)—higher than the average for "all races"—followed by infants of Mexican (5.0), Central and South American (4.5), and Cuban (4.1). Since 2005, the infant mortality rate has declined by 19 percent for all mothers and 10 percent for Hispanic mothers. Across Hispanic subgroups, the rate dropped 26 percent for Puerto Rican women, 10 percent for Mexican women, 6 percent for Cuban women, and 3 percent for Central and South American women. The graph shows similar rates between Mexicans and Central Americans relative to non-Hispanic White mothers. This evidence supports the Hispanic health paradox given that Hispanics have lower socioeconomic status relative to non-Hispanic Whites.

Leading Causes of Death

Patterns in the causes of death—both between Hispanics and other Americans, as well as between Hispanic subgroups—may help to explain the health paradox. We report mortality rates by cause-of-death in Table 1.

Heart disease and cancer are the two leading causes of death for all population groups: non-Hispanic White, non-Hispanic Black, and Hispanics. Interestingly, some elements of the cause-of-death data seem to sharpen the Hispanic health paradox. For example, while Hispanics have the highest life expectancy at birth and the lowest death rates of all populations, they also have some of the highest disease-specific death rates. Hispanics have higher age-adjusted death rates than the non-Hispanic White population for diabetes, kidney disease, and chronic liver and cirrhosis, and higher age-adjusted death rates than the non-Hispanic Black population due to chronic liver and cirrhosis. These differences only come to light when using the age-adjusted rates. The observed differences in death rates due to diabetes, liver disease, and kidney disease disappear when using unadjusted rates (with additional details in Table A2 in the Appendix).

While crude death rates increase with older populations, age-adjusted rates are constructed based on assumptions of a baseline population distribution. To address these problems, Table 1 shows unadjusted rates by the leading cause of death for different age and Hispanic subgroups. Among Hispanic subgroups, we find significant differences in leading causes of death. Cubans and Puerto Ricans have higher death rates of heart disease, cancer, and Alzheimer's disease. Mexicans aged 55 and above have higher death rates associated with diabetes, kidney disease, and chronic liver disease and cirrhosis than any other Hispanic subgroup. Cubans are older than the other Hispanic subgroups, which helps to explain Cubans higher cause-of-death rates for diseases that predominantly affect older adults like Alzheimer's and heart disease.

Morbidity

Despite the observed advantage of the Hispanic population in aggregated mortality rates, other measures of health among Hispanics offer a mixed picture. We consider the most common morbidities discussed in the literature that directly

Table 1
Leading Causes of Death-Unadjusted Death Rates

	Cubans	*Puerto Ricans*	*Hispanics*	*Mexicans*	*South Americans*	*Central Americans*
Heart disease						
All	263.8	133.6	96.5	79.3	67.0	51.4
35–54	32.8	49.8	37.4	37.0	13.4	26.3
55–74	239.1	290.5	236.9	239.4	87.2	152.6
75+	2,283.0	2,039.7	1,840.5	1,719.1	1,248.8	1,371.3
Cancer						
All	148.2	84.6	71.2	60.6	71.3	44.1
35–54	34.2	41.9	39.7	39.2	27.1	32.3
55–74	284.5	272.3	251.3	243.8	178.1	172.5
75+	872.9	848.1	869.1	843.3	780.7	711.2
Alzheimer's disease						
All	41.4	17.8	13.6	11.5	10.2	5.0
35–54	0.2	0.2	0.1	0.1	0.0	0.1
55–74	8.5	14.4	8.1	7.9	2.8	3.3
75+	442.1	419.7	424.5	447.4	260.7	251.6
Diabetes						
All	23.0	20.3	16.8	16.7	6.2	9.1
35–54	4.4	10.8	8.6	9.3	1.5	5.0
55–74	37.3	59.3	57.8	66.9	12.1	36.8
75+	155.3	223.2	228.9	267.1	95.4	180.1
Chronic liver and cirrhosis						
All	7.1	9.5	11.4	11.9	4.0	8.1
35–54	3.5	8.5	13.8	15.2	2.4	11.5
55–74	18.2	38.0	43.6	51.3	10.9	28.6
75+	25.8	37.9	55.1	69.2	33.9	63.1
Kidney disease						
All	8.8	9.5	7.4	7.5	3.6	3.6
35–54	0.6	3.6	2.9	3.2	0.5	1.5
55–74	8.5	26.0	22.6	26.5	5.9	14.0
75+	77.0	122.0	121.6	145.0	62.2	78.9

Source: Authors' calculations.
Notes: Mortality rates (deaths per 100,000 population) are calculated using mortality data from the National Vital Statistics System in 2019 and population from the 2019 American Community Survey. For explicitly age-adjusted cause-of-death rates, see Table A2 in the online Appendix.

connect to our previous measures of leading causes of death. Advantages and disadvantages on death rates are likely related to risk factors reflected in morbidity rates. Morbidity is measured as the proportion of individuals within a group with a particular health condition. We measure morbidity using nationally representative data from the National Health Interview Survey from 2016 to 2019, as harmonized by Integrated Public Use Microdata Series (IPUMS) where respondents self-report their ethnicity, country of birth, and medical conditions.

We concentrate our analysis on all adults (age > 17) who were "ever" diagnosed with a particular condition during this time period. We combine four years of data to increase the sample size. The larger sample size provides us with statistical power to explore disaggregated Hispanic groups based on nativity and ancestry. Additionally, we can isolate within group effects from observed gender/age variation.[7] We estimate prevalence rate differences between each Hispanic group relative to non-Hispanic Whites conditional on age, sex, and survey year cohort.

Overall, we find evidence consistent with earlier studies (Markides and Coreil 1986; Sorlie et al. 1993; Abraído-Lanza et al. 1999; Hummer et al. 2000). Hispanics display advantages in cancer (–4 percentage points), cervical cancer (–2 percentage points), and coronary heart disease (–0.5 percentage points). The cardiovascular disease indicators of high blood pressure and hypertension, which are normally positively correlated, give mixed results (–1 percentage point and 1 percentage point, respectively). Conversely, prevalence rates for diabetes, kidney failure, and chronic liver disease are higher for Hispanics than non-Hispanic Whites. On average, Hispanics are more likely to have ever been diagnosed with diabetes (4 percentage points), kidney failure (1 percentage point), and chronic liver disease (0.5 percentage points) than non-Hispanic Whites.

When we disaggregate Hispanics by ancestry and nativity, the differences across Hispanic subgroups become more apparent (where "Island" refers to those born in Puerto Rico but now living on the mainland). Figure 3 highlights the differences in proportions between Hispanics and non-Hispanic Whites across key morbidities and separates Hispanics across nativity/ancestry.[8] The comparison is not only within Hispanic ethnicity but also within country/ancestry identification and place of birth, allowing us to highlight important differences. For example, Mexican immigrants have substantially lower rates of hypertension than Mexican Americans. Hispanics have a higher diagnosis rate for chronic illnesses like diabetes, hypertension, kidney failure, or chronic liver disease than non-Hispanic Whites. This rate is driven by US-born Hispanics rather than by the Immigrant/Island group. For all Hispanic groups, the estimated prevalence rate differences among foreign/island-born Hispanics are lower. Consistent with Young and Hopkins (2014) regarding Hispanics advantage on cancer morbidity rates, this advantage persists throughout all the disaggregation exercises.

Obesity

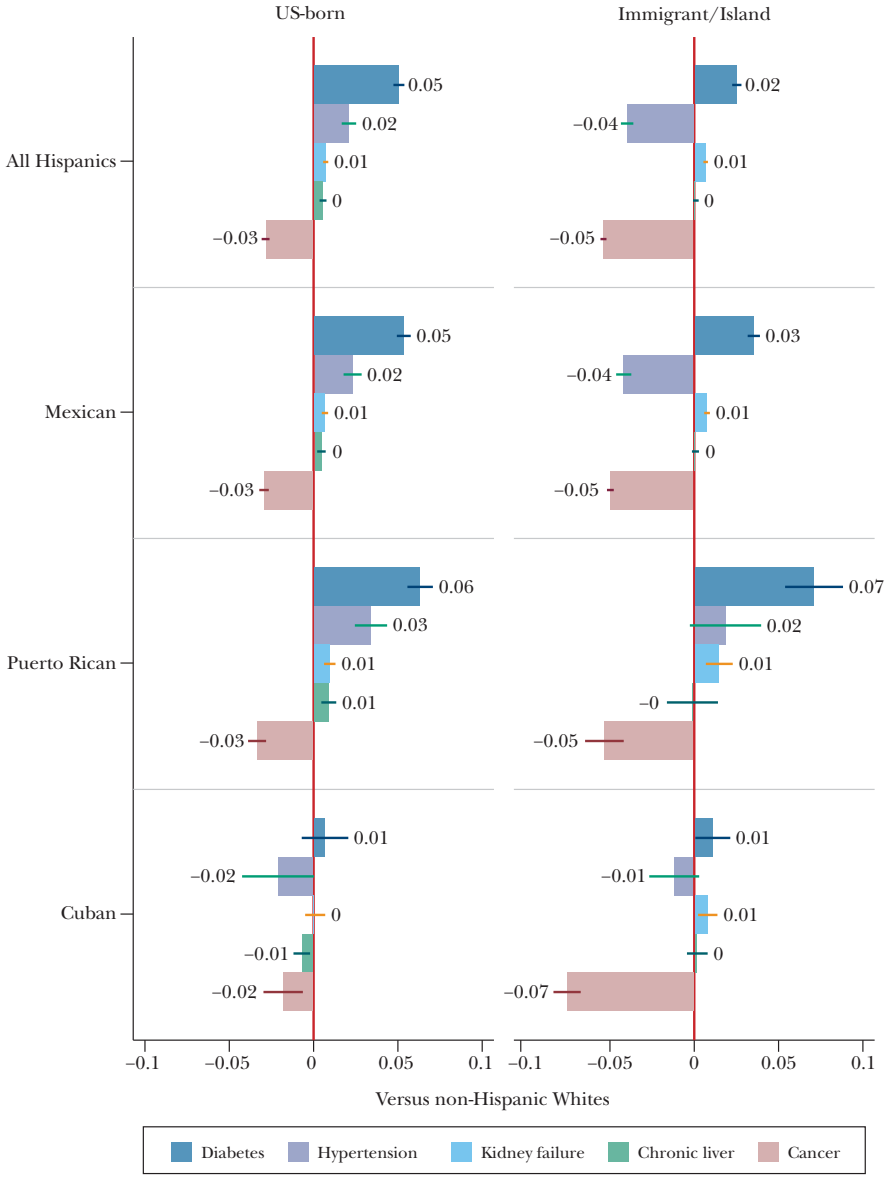
The analysis of diagnosed medical conditions allowed us to illustrate the potential health advantages and disadvantages among Hispanics framed within the leading mortality indicators and the Hispanic paradox. However, another health measure that has researchers' attention is obesity rates among Hispanics. Obesity

[7] Figures A1–A5 in the Appendix show a breakdown of the estimates by gender and age group.

[8] Due to confidentiality issues, ancestry information for respondents selecting a Central or South American country are aggregated to the regional level, limiting our ability to disaggregate even further this group.

Figure 3

Difference in the Likelihood of Ever Being Diagnosed with Condition Relative to Non-Hispanic White



Source: Authors' calculations using 2006–2019 NHIS-IPUMS data.

Notes: The results are the estimated differences in the likelihood of ever being diagnosed with a condition between the identified group and non-Hispanic Whites after controlling for age, sex, and survey year fixed effects. Estimated average diagnostic rates for diabetes (14.24 percent), hypertension (46.72 percent), kidney failure (3.53 percent), chronic liver condition (2.08 percent), and cancer (9.1 percent) are the baseline averages. Values result from linear regressions of ever being diagnosed with the corresponding condition controlled by age and gender with Non-Hispanic Whites as the reference group. Lines represent the confidence intervals. Individuals are classified within the country's ancestry/origin group they self-identified. All Hispanics aggregates all individuals who self-identified as Hispanics in the survey.

is a risk factor that helps to explain the development of other conditions such as cardiovascular disease and stroke. On average, Hispanics have a lower obesity rate compared to non-Hispanic Whites. Also, the age-adjusted percentage of Hispanics that are obese is 45 percent relative to 42 percent for Whites (Hales et al. 2020). A few researchers have found that the likelihood of obesity is highest among Mexicans and Puerto Ricans (Isasi et al. 2015). However, obesity appears to be a growing problem in Hispanic communities. Recent immigrants have lower rates of obesity, but obesity rates increase as time spent in the US increases (Ai, Appel, and Lee 2018).

Mental Health

The majority of our illustrations of the Hispanic health paradox have centered on physical health, but mental health deserves attention too. The evidence is mixed for mental health, but in general, Hispanics, both immigrants and natives, have a lower prevalence of mental health issues, particularly among Puerto Ricans and Cubans (Alarcón et al. 2016). Hispanics have lower rates of depression and suicide than non-Hispanic Whites. Immigrant Hispanics are less likely to report anxiety, depression, or other disorders compared to Hispanic Americans (Vega et al. 2004), but this reverses the longer the immigrant remains in the United States (Cook et al. 2009).

Leading Explanations

The Hispanic health paradox remains an unsolved puzzle. Here, we explore a range of possible explanations, seeking to describe what research has been done and some promising directions for future research.

Demographics and Socioeconomic Differences

Demographic factors can partially explain the original paradoxical findings of Hispanic health statistics. The different age and gender distributions between Hispanics and non-Hispanic Whites have accounted for some of the mortality and life expectancy advantages. However, they do not fully account for the differences. For instance, in the case of lower infant mortality among Hispanics, some of the paradox is explained by younger maternal age among Hispanic mothers, especially Hispanic immigrant mothers (Hummer et al. 1999). Infant mortality increases at older maternal ages (Powers 2013). While different factors can affect infant mortality, socioeconomic disadvantages have been strongly and consistently associated with higher infant mortality rates—except for the case of infants born to mothers of Mexican origin (Elder, Goddeeris, and Haider 2016).

Sample Selection Bias

There are two main reasons why Hispanics might not be accurately depicted in data sources, in a way that can cause estimates of their health to be biased. First, the "healthy immigrant effect" refers to the pattern that in any host country, the immigrant population may be healthier on average than the non-immigrant population.

In general, healthier individuals are more willing to pay the cost of immigration. As a result, recent immigrants are individuals who are positively selected on health and thus have better outcomes when compared to the general US population. However, subsequent generations of these immigrants show regression to the mean as their children's health outcomes tend towards those of non-Hispanic Whites (García-Pérez 2016).

Second, the salmon bias hypothesis is a negative selection effect that refers to a sample selection bias resulting from return migration. Pablos-Méndez (1994, p. 1237) pointed out that "many Hispanics return to their country of birth when they retire, become severely ill, or simply after a temporary job." He referred to this pattern as "salmon bias," "highlighting the compulsion to die in one's birthplace." The deaths of those who return to their country of origin will not be recorded in US mortality statistics: as Pablos-Méndez wrote, "[S]ome individuals are rendered statistically immortal." As a result, the immigrants that remain in the US will tend to be younger and healthier than those who return. Among the other reasons to return to the country of origin, researchers find a lower cost of living, the presence of family members, and lower return migration costs (Arenas et al. 2015). In early studies, Jasso and Rosenzweig (1982) found that all immigrant Hispanics, except for Cubans, have large emigration rates. Conversely, Abraído-Lanza et al. (1999) find evidence rejecting that the salmon bias hypothesis explains the Hispanic health paradox without ruling out some role for selective migration.

These sources of bias can partially explain why observed Hispanic health advantages appear uneven and are not fully generalizable across Hispanic subgroups. For example, we observe more elderly Cubans, relative to other Hispanic groups, in part because the costs of return migration to Cuba have been nearly prohibitive due to political forces. We also observe higher prevalence of elderly-related conditions among Cubans. Conversely, undocumented individuals (largely connected to Mexican migrants) experience a much higher cost of obtaining medical services if they remain in the United States, and thus have an additional incentive to return to their origin country compared to documented immigrants. We observe higher prevalence of chronic conditions among Mexicans, yet lower among immigrant Mexicans. Notwithstanding, the experience of Puerto Ricans can be especially relevant to unraveling the Hispanic health paradox. Puerto Ricans are not immigrants: they are eligible for all US health care programs like Medicare and Medicaid. There are potentially other care access issues affecting this group's differential health outcomes.

The healthy immigrant effect and the salmon bias hypothesis are not mutually exclusive: indeed, they would tend to reinforce each other in supporting the Hispanic health paradox. Several studies have sought to disentangle these two potential sources of selection. Riosmena, Wong, and Palloni (2013) combine data from the Mexican Health and Aging Study in Mexico and the US National Health Interview Survey and find evidence for the existence of both healthy immigrant bias and salmon bias, but also find that they are only a partial explanation for the Hispanic health paradox.

In our own analysis, using age-adjusted death rates significantly reduced the Hispanic health advantage, suggesting the salmon effect has some bite. When we controlled for ancestral country and place of birth, our measures of morbidity described immigrant Hispanics as healthier on average than native Hispanics. Even when considering obesity and mental health rates, the healthy immigrant effect persisted.

However, these two biases alone cannot explain the paradox. One would assume that in the absence of these biases, foreign-born individuals will look similar to their native counterparts. However, differential access to health care is likely to remain due to immigration status, residential location, lack of insurance, and language barriers. If anything, the presence of these barriers suggests that the underlying size of some of the described health advantages—net of these barriers—may be underestimated.

Measurement Error

Survey data of self-reported status and outcomes are always prone to measurement error. In our case, self-reported health outcomes, health status, and Hispanic/race identity are areas of concern (Chatterji, Joo, and Lahiri 2012). Collecting, recording, reporting, and counting deaths and births among Mexican Americans, especially around US border counties, can create accounting issues for the aggregated rates (Markides and Eschbach 2005). Even the question of citizenship has flaws, with some Puerto Ricans appearing as noncitizens in the American Community Survey (Brown et al. 2019). Here we discuss three factors that can create a bias within the Hispanic health paradox.

First, in order to avoid detection, undocumented immigrants may be less likely to answer surveys or to use health care, a fact sometimes known as the "chilling effect." Even when health care is used, undocumented immigrants may instead focus on only their immediate ailments, thereby never documenting a broader diagnosis. As a result, survey questions focused on Hispanics living in the United States or the use of medical records may both underestimate health issues for this population. This chilling effect can have an externality even among documented individuals who fear an undocumented family member may become exposed. Alsan and Yang (2022) find that Hispanic citizens reduce their participation in the Supplemental Nutrition Assistance and Social Security Income programs when immigration enforcement intensifies.

Second, researchers point to "ethnic attrition," the tendency of second and later generations of Hispanic immigrants to fail to self-identify as Hispanics, resulting in a downward bias in the estimated health of children of immigrants as a result of assimilation (Antman, Duncan, and Trejo 2020). For example, approximately half of all fourth-generation Hispanics still identify as Hispanic (Lopez, Krogstad, and Passel 2022). As the rate of ethnic attrition increases, aggregate values of vital statistics and health outcomes become more immigrant-centric. The direction of the bias will depend on the health status of those Hispanics who stop identifying. If healthier individuals are more likely not to identify, then the observed health

advantages should diminish. However, if sicker individuals stop identifying, health advantages will only increase.

Third, health care usage among immigrants has been tied to length of time in the United States. A shorter tenure implies less health care usage as immigrants may have trouble navigating the US healthcare system. A by-product of this behavior is that the children of immigrants may also have a reduction in access to and usage of health care, resulting in an underreporting of health outcomes (García-Pérez 2013; 2016).

Cultural and Lifestyle Differences

Cultural and social factors could potentially protect individuals from developing certain negative health outcomes. These factors provide an informal support mechanism for care. Culture can shape an individual's risk and lifestyle behaviors. Therefore, if Hispanics differ from other groups in categories concerning smoking, alcohol consumption, risky behavior, and food consumption, these community-constructed individual behaviors could result in a collective gain in terms of health outcomes. Strong social and family ties are associated with reductions in stress and anxiety, but community factors can also reverse that positive relation, such as discrimination and language barriers (Alegria et al. 2007). Eschback et al. (2004) find evidence of a "barrio neighborhood advantage" to explain low adult mortality among Hispanics living in immigrant neighborhoods. However, Palloni and Arias (2004) find no evidence of cultural/social factors, such as marital status and segregation index, to explain the advantages in adult mortality rates.

Smoking and alcohol consumption has consistently been connected to lower risk factors for developing conditions such as cancer and cardiovascular disease, lower mortality, and higher life expectancy among Hispanics. Smoking and alcohol consumption habits are often influenced by social interactions. In the case of low infant mortality, the literature emphasizes the cultural aspect of caring for expecting mothers in the Mexican American community. One potential explanation for the paradox is the differential smoking and drinking rates of Hispanic immigrants versus Hispanic Americans. Immigrants are less likely to drink or smoke, which could contribute to better infant outcomes.

Hispanics daily smoking rate is 8 percent, which is lower than that of non-Hispanic Whites (Cornelius et al. 2022). Puerto Ricans and Cubans are more likely to smoke compared to Mexicans, Dominicans, and Central Americans (Martell, Garrett, and Caraballo 2016; Kaplan et al. 2014). Hispanic immigrants display positive selection in that they have lower smoking rates than individuals in their home country and Hispanics in the United States (Bosdriesz et al. 2013).

Similarly, Hispanics are less likely to drink alcohol when compared to Whites. Seventy percent of White Americans reported having one drink in the past year compared to 54 percent of Hispanics (National Institute on Alcohol Abuse and Alcoholism 2021). However, Hispanics are more likely to binge drink than Whites (42 percent versus 32 percent for Hispanics and White drinkers, respectively). Puerto Ricans have the highest percentage of drinkers, binge drinkers, and

individuals with alcohol dependence, while Cubans report the lowest percentage across all of these categories.

The Role of Health Insurance and Usage in the Hispanic Health Paradox

Health Insurance

We believe that the potential role of health insurance in the Hispanic health paradox has been understudied. Aggregate statistics suggest that Hispanics tend to have lower-than-average health insurance rates and health care use. Such patterns might potentially affect the Hispanic health paradox in two ways.

First, individuals with access to insurance coverage or greater use of health care may become more aware of their health and more likely to report specific health conditions in a survey. Second, to the extent that differential access to health insurance also leads to differences in the usage of health care, it may also lead to differences in recorded health outcomes. The lack of health insurance may lead to a greater degree of survey nonresponse for certain health conditions, leading some to believe that Hispanics are simply healthier. The lack of insurance leading to less health care usage would imply that administrative claims data would also underreport certain health outcomes. To go one step further, it might imply that if Hispanics had equal rates of health insurance and health care usage, the Hispanic health paradox might be even larger.

Which of these effects is likely to dominate? One approach would be to look at trends over time. For example, if increased health insurance coverage for Hispanics leads to worse reported health statistics, it would be consistent with insurance leading to heightened awareness and reporting of health problems. Conversely, if increased health insurance coverage for Hispanics leads to improved health statistics, it would imply the Hispanic health paradox is stronger than previously believed.

In this subsection, we discuss patterns of health insurance coverage for Hispanics in the last 15 years or so. In the next subsection, we consider patterns of health care usage for Hispanics. In both discussions, we sketch the fact base in these areas and offer some preliminary thoughts, while emphasizing a need for additional research.

Using data from the American Community Survey via the Integrated Public Use Microdata Series (IPUMS), we can identify recent patterns in health insurance coverage by race, ethnicity, and citizenship. Health insurance coverage among Hispanics increased from 69.1 percent in 2008 to 82.6 percent in 2020. The difference in coverage rates between Hispanics and non-Hispanic Whites has decreased over this same time period from 20.6 to 11.2 percentage points. The improvements in health insurance rates are largely attributable to the Patient Protection and Affordable Care Act of 2010. Both private and public health insurance rates for Hispanics increased by approximately 7 percentage points each. Medicare coverage

Table 2

Percent of Any Health Insurance Coverage by Ancestral Heritage and Citizenship in 2020

	Citizen (percent)	Non-citizen (percent)
Non-Hispanic Whites	93.9	87.9
Hispanics	88.2	57.3
Mexican	86.6	53.9
Puerto Rican	92.0	—
Cuban	90.9	71.4
Central American	88.1	49.0
South American	90.7	70.4
Other	90.3	73.0

Source: Author calculations using the 2020 American Community Survey.
Note: Less than 2 percent of Puerto Ricans report being noncitizens, but since all Puerto Ricans are US citizens, this percentage is not reported in the table.

for Hispanics rose by only 2.3 percentage points, which is the smallest rise among all major race and ethnicity groups.

Table 2 reports the percentage of individuals with any type of health insurance by ancestral heritage and citizenship in 2020. Among Hispanic citizens, the percentage of people with any form of health insurance is comparable to non-Hispanic citizens, ranging from 86.6 percent to 92 percent. However, the range is much wider among noncitizens, with values between 49 percent and 84.4 percent. These values are all lower than the rate of insurance among noncitizen/non-Hispanics, at 86.3 percent.

Disaggregating Hispanics into countries of ancestral heritage, we observe coverage rates ranging from a low of 80.6 percent for those of Mexican heritage and a high of 91.9 percent for those of Puerto Rican heritage in 2020. When we further separate these groups between US citizens and noncitizens, the differences become starker. Hispanic citizens experienced an increase in coverage from 80 percent to 88.2 percent, while Hispanic noncitizens experienced an even larger increase from 39.7 percent to 57.4 percent from 2008 to 2020. Despite the increase in coverage among noncitizens, the average health insurance gap between citizens and non-citizen Hispanics is 30.8 percentage points.

Public insurance coverage for Hispanics decreased from 38.7 percent in 2016 to 36 percent in 2020, with Medicaid coverage falling by 3.7 percentage points. These decreases in public insurance rates appear to be offset by a 4.6 percentage point increase in private insurance coverage. Yue, Rasmussen, and Ponce (2018) find that Medicaid expansion policies in the aftermath of the 2010 Affordable Care Act were relatively weak among Hispanics. Moreover, these Medicaid expansion policies were not found to have a statistically significant effect on health insurance coverage or on health care access measured by having a regular doctor and frequency of flu shots. Even more puzzling is that the health insurance coverage gap between Hispanics

and non-Hispanic White people is larger in states that have expanded Medicaid than in those that have not.

There are two potential reasons for these puzzling results. First, large Hispanic populations in Florida and Texas—states that did not expand Medicaid coverage—decrease the potential benefit of Medicaid expansion to Hispanics. Approximately 35 percent of the Hispanic population lives in non-Medicaid expansion states. Second, increased immigration enforcement may have caused a "chilling effect" on healthcare usage in states with more Immigration and Customs Enforcement (ICE) activity. One (admittedly imperfect) proxy for the intensity of immigration enforcement is I-247 "detainer requests." An I-247 request is a notice from ICE to local law enforcement that ICE intends to assume custody of an individual currently being held by local law enforcement. Watson (2014) finds an 8.7 percent decline in Medicaid participation among children of noncitizen parents after a 1 percent increase in I-247 detainer requests. Friedman and Venkataramani (2021) find that health care usage among Hispanics decreases after a one standard deviation increase in I-247 requests per capita, even for patients with chronic conditions such as diabetes, but there is no difference for the non-Hispanic White population.

Economists can explore if changes to health insurance access—through policies such as the Patient Protection and Affordable Care Act of 2010 or through Medicaid expansion—affect the Hispanic health paradox.[9] The accessibility of affordable health insurance could explain some of the perceived advantages of noncitizen Hispanics over citizen Hispanics. A lack of health insurance could imply fewer visits to the doctor's office. These fewer visits could mean that important information about health may never be recorded. Disease prevalence rates are likely measured with error in the uninsured community as only those with severe cases will seek care. Less severe cases are more likely to go undocumented. Given the large gap in insurance coverage between citizens and noncitizens for some Hispanic groups, we would expect larger changes in healthcare usage among Central American and Mexicans relative to Cubans, Puerto Ricans, and South Americans, who have a smaller gap.

Health Care Usage

The presence of a diagnosis, disease awareness, and the usage of health care are intertwined. Using data from the National Health Interview Survey, we calculate two measures of health care usage intent: whether one has a usual place of care and whether the usual place of care is an emergency service. We measure actual health care usage by responses to having visited the doctor in the past two years and having visited an emergency room in the past twelve months.

[9] The Deferred Action for Childhood Arrivals (DACA) provisions of US immigration policy have allowed some individuals who arrived in the United States as children and without legal authorization to participate in the state-run health insurance exchanges set up under the Patient Protection and Affordable Care Act of 2010, but they are not eligible for the subsidies provided to US citizens.

We compare health care usage differences between Hispanic groups and non-Hispanic Whites, controlling for age, sex, and survey year variations.[10] We find that 56 percent of Hispanics overall have no usual place of care. However, Hispanics as a group are 7 percentage points more likely to lack a usual place of care than non-Hispanic Whites. The lack of a usual place of care among all the Hispanic groups suggests the possibility that a large proportion of this population is missing preventive care, either because of limited access to quality health care or overall barriers to access to care. In particular, Mexicans, regardless of place of birth, are more likely to lack a usual place of care, which may reflect a lack of access to healthcare, especially primary care, in the areas where these populations traditionally reside.

On average, Hispanics are more likely than non-Hispanic Whites to identify emergency rooms as their usual place of care (6 percent versus 4 percent, respectively). Foreign/island-born Hispanics have double the rate of non-Hispanic Whites. The differential rates across Hispanic subgroups compared to non-Hispanic Whites vary significantly across nativity and ancestry. For Cubans and Mexican immigrants, the differential rates are higher than their native counterparts, but for Puerto Ricans, the differential rate is only significant and positive among those born on the mainland. Foreign-born Mexicans are 2 percentage points more likely to use the emergency room as a usual place of care compared to US-born Mexicans. However, island-born Puerto Ricans are 9 percentage points less likely compared to Puerto Ricans born on the mainland. The inefficient use of emergency services relative to a traditional doctor's office are well known, including higher medical expenses in health care and higher out-of-pocket expenses for patients, and the possibility that ailments may worsen before treatment (DuBard and Massing 2007; Tarraf, Vega, and González 2014; Basu Roy, Olsen, and Tseng 2020; Zhao and Nianogo 2022).

The average rate of visiting a doctor in the last two years for the entire population is 86 percent. All Hispanic groups are less likely to have visited the doctor in the last two years. The foreign-born Mexicans, Cubans, and other Hispanics lead the estimated differences by –8 percentage points, –3 percentage points, and –4 percentage points, respectively, compared to non-Hispanic Whites.

The average rate of emergency room visits for the entire population is 20 percent. The results do not indicate high use of emergency room services by most Hispanic groups. Despite a plausible intuition that this group might overuse emergency care rooms due to a lack of preventive care, all Hispanic immigrant groups have a lower probability of visiting emergency rooms than non-Hispanic Whites.

As we alluded to before, health insurance has a significant effect on health usage behavior. Simply controlling for health insurance reduces the gap between Hispanics and non-Hispanic Whites by about half across all categories, except for emergency room visits. Additionally, we consider the within-group differences in health care usage conditional on having health insurance. We see the greatest differences among Cubans, where the difference in having a normal place of care is

[10] See Appendix for detailed results (Figures A6-A7).

13 percentage points higher for Cuban Americans and 27 percentage points higher for immigrant Cubans.[11]

There are several takeaways from the differences between health care usage and mortality. First, the low rate of infant mortality among Hispanics is even more impressive given the lack of health care usage. Second, the lack of health care usage, particularly preventive care, could explain the higher morbidity rates of hypertension, diabetes, and liver diseases. Additionally, the lack of health care not only affects improvements in health, but also prevents disease awareness.

We have demonstrated throughout this essay that using Hispanics as an aggregate monolithic group hides variations in health outcomes by subgroup ancestry. We have shown that nativity can play a role in the paradox, both through ethnic attrition and assimilation, to explain why the healthy immigrant effect diminishes in future generations. We have provided evidence throughout this essay that "salmon bias" could be contributing to the perceived advantages in Hispanic mortality by comparing unadjusted and age-adjusted death rates. We have explained that access to health insurance access and health care not only affects the health of individuals directly, but ultimately affects how and if the measures we use to account for the paradox are ever recorded.

Concluding Remarks

The Hispanic health paradox remains a ripe subject for further research and probably does not have a single unique cause. Instead, these pieces of the puzzle may affect Hispanic subgroups differently according to their birthplace, place of ancestry, status as documented or undocumented immigrants, length of time residing in the country, geographic residential location in the United States, and age/gender/socioeconomic compositions. Broad statements about Hispanics as a group often do not translate into better comparable health outcomes among all Hispanics (Jerant, Arellanes, and Franks 2008). Hispanics' mortality rates heterogeneity is expected to reflect the differences in health outcomes as well as access and usage of healthcare, which are further accentuated when foreign-born/island status is considered (Borrell and Crawford 2009). Moreover, the composition of Hispanics has changed dramatically in recent decades: it was an immigrant-dominated group prior to 1990, but has been a citizen-dominated group since 2010. Therefore, the membership in this group is not time-invariant.

In the context of health disparities, the fact that the Hispanic subgroups do not have consistent patterns calls for more research. In the discussion of the paradox, it is necessary to identify the potential mechanisms behind lower health care usage, differential health outcomes, and preventable costs, especially among elderly Hispanics and those suffering chronic diseases. As this research continues,

[11] See Appendix for full results (Table A.3).

we suspect that three factors will play an important role. One is understanding the interaction of health insurance, health care usage, and preventive care, especially regarding the effects among elderly Hispanics and those suffering from chronic diseases. Next, when analyzing health care usage and costs for Hispanics, it will be important to disaggregate Hispanic subgroups by place of birth and age profile. Third, measurement error is likely playing a larger role than previously suspected. Economists could develop behavioral/empirical models to address external barriers to care (including living in rural areas), self-selection when seeking health care, and the presence of measurement error in health data.

Ultimately, the Hispanic health paradox offers a starting point for a deeper examination of what leads to differences in health outcomes—and thus a fuller understanding of how to address the underlying health issues more directly and how these issues would worsen after the disparate effects of the COVID-19 pandemic.

References

Abraído-Lanza, Ana F., Bruce P. Dohrenwend, Daisy S. Ng-Mak, and J. Blake Turner. 1999. "The Latino Mortality Paradox: A Test of the 'Salmon Bias' and Healthy Migrant Hypotheses." *American Journal of Public Health* 89 (10): 1543–48.

Ai, Amy L., Hoa B. Appel and Jungup Lee. 2018. "Acculturation Factors Related to Obesity of Latino American Men Nationwide." *American Journal of Men's Health* 12 (5): 1421–30.

Alarcón, Renato D., Amrita Parekh, Milton L. Wainberg, Cristiane S. Duarte, Ricardo Araya, and María A. Oquendo. 2016. "Hispanic Immigrants in the USA: Social and Mental Health Perspectives." *The Lancet Psychiatry* 3 (9): 860–70.

Alegría, Margarita, Norah Mulvaney-Day, Maria Torres, Antonio Polo, Zhun Cao, and Glorisa Canino. 2007. "Prevalence of Psychiatric Disorders across Latino Subgroups in the United States." *American Journal of Public Health* 97 (1): 68–75.

Ali, Jennifer. 2002. "Mental Health of Canada's Immigrants." *Supplement to Health Reports,* Vol. 13. Ottawa: Health Canada.

Alsan, Marcella and Crystal S. Yang. 2022. "Fear and the Safety Net: Evidence from Secure Communities." *The Review of Economics and Statistics* 1–45. https://doi.org/10.1162/rest_a_01250.

Amuedo-Dorantes, Catalina, Brandyn Churchill, and Yang Song. 2022. "Immigration Enforcement and Infant Health." *American Journal of Health Economics* 8 (3): 323–58.

Antman, Francisca M., Brian Duncan, and Stephen J. Trejo. 2020. "Ethnic Attrition, Assimilation, and the Measured Health Outcomes of Mexican Americans." *Journal of Population Economics* 33 (4): 1499–1522.

Arenas, Erika, Noreen Goldman, Anne R. Pebley, and Graciela Teruel. 2015. "Return Migration to Mexico: Does Health Matter?" *Demography* 52 (6): 1853–68.

Arias, Elizabeth. 2010. "United States Life Tables by Hispanic Origin." *Vital and Health Statistics. Series 2, Data Evaluation and Methods Research* (152): 1–33.

Bacon, Emily, Fernando Riosmena, and Richard G. Rogers. 2017. "Does the Hispanic Health Advantage Extend to Better Management of Hypertension? The Role of Socioeconomic Status, Sociobehavioral Factors, and Health Care Access." *Biodemography and Social Biology* 63 (3): 262–77.

Basu Roy, Subhasree, Reed Neil Olsen, and Huikuan Tseng. 2020. "Do Hispanic Immigrants Spend Less on Medical Care? Implications of the Hispanic Health Paradox." *Applied Economics* 52 (36): 3951–64.

Blackwell, Debra L., Jacqueline W. Lucas, and Tainya C. Clarke. 2014. "Summary Health Statistics for U.S. Adults: National Health Interview Survey, 2012." *Vital Health Statistics. Series 10, Data from the National Health Survey* 260: Table 25.

Blewett, Lynn A., Julia A. Rivera, Miriam L. King, Kari C.W. Williams, Natalie Del Ponte, and Pat Convey. "IPUMS Health Surveys: National Health Interviews Survey, Version 8.1." IPUMS. https://doi.org/10.18128/D070.V8.1. (accessed February 11, 2022).

Borrell, Luisa N., and Natalie D. Crawford. 2009. "All-Cause Mortality Among Hispanics in the United States: Exploring Heterogeneity by Nativity Status, Country of Origin, and Race in the National Health Interview Survey-Linked Mortality Files." *Annals of Epidemiology* 19 (5): 336–43.

Bosdriesz, Jizzo R., Nienke Lichthart, Margot I. Witvliet, Wim B. Busschers, Karien Stronks, and Anton E. Kunst. 2013. "Smoking Prevalence among Migrants in the US Compared to the US-Born and the Population in Countries of Origin." *PloS One*, 8 (3): e58654.

Brown, J. David, Misty L. Heggeness, Suzanne M. Dorinski, Lawrence Warren, and Moises Yi. 2019. "Predicting the Effect of Adding a Citizenship Question to the 2020 Census." *Demography* 56 (4): 1173–94.

Chatterji, Pinka, Heesoo Joo, and Kajal Lahiri. 2012. "Beware of Being Unaware: Racial/Ethnic Disparities in Chronic Illness in the USA." *Health Economics* 21 (9): 1040–60.

Constant, Amelie F., Teresa García-Muñoz, Shoshana Neuman, and Tzahi Neuman. 2018. "A 'Healthy Immigrant Effect' or a 'Sick Immigrant Effect'? Selection and Policies Matter." *The European Journal of Health Economic*s 19 (1): 103–121.

Cook, Benjamin, Margarita Alegría, Julia Y. Lin, and Jing Guo, 2009. "Pathways and Correlates Connecting Latinos' Mental Health with Exposure to the United States." *American Journal of Public Health* 99 (12), 2247–54.

Cornelius Monica E., Caitlin G. Loretan, Teresa W. Wang, Ahmed Jamal, David M. Homa. 2022. "Tobacco Product Use Among Adults—United States, 2020." *Morbidity and Mortality Weekly Report* 71 (11): 397–405.

DuBard, C. Annette, and Mark W. Massing. 2007. "Trends in Emergency Medicaid Expenditures for Recent and Undocumented Immigrants." *JAMA* 297 (10): 1085–92.

Eschbach, Karl, Glenn V. Ostir, Kushang V. Patel, Kyriakos S. Markides, and James S. Goodwin. 2004. "Neighborhood Context and Mortality among Older Mexican Americans: Is There a Barrio Advantage?" *American Journal of Public Health* 94 (10): 1807–12.

Elder, Todd E., John H. Goddeeris, and Steven J. Haider. 2016. "Racial and Ethnic Infant Mortality Gaps and the Role of Socio-Economic Status." *Labour Economics* 43: 42–54.

Ely, Danielle M., and Anne K. Driscoll. 2021. "Infant Mortality in the United States, 2019: Data from the Period Linked Birth/Infant Death File." *National Vital Statistics Reports* 70 (14): 1–18.

Fernandez, José, Mónica Garcia-Pérez, and Sandra Orozco-Aleman. 2023. "Replication data for: Unraveling the Hispanic Health Paradox." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E182763V1.

Friedman, Abigail S., and Atheendar S. Venkataramani. 2021. "Chilling Effects: US Immigration Enforcement and Health Care Seeking among Hispanic Adults: Study Examines the Effects of US Immigration Enforcement and Health Care Seeking among Hispanic Adults." *Health Affairs* 40 (7): 1056–65.

García-Pérez, Mónica. 2013. "Health Care Usage and Health Status of Immigrant Children: The Effects of Nativity versus Citizenship." *American Economic Review* 103 (3): 412–17.

García-Pérez, Mónica. 2016. "Converging to American: Healthy Immigrant Effect in Children of Immigrants." *American Economic Review* 106 (5): 461–66.

Goel, Mita Sanghavi, Ellen P. McCarthy, Russell S. Phillips, and Christina C. Wee. 2004. "Obesity among US Immigrant Subgroups by Duration of Residence." *JAMA* 292 (23): 2860–7.

Hales, Craig M., Margaret D. Carroll, Cheryl D. Fryar, and Cynthia L. Ogden, 2020. "Prevalence of Obesity and Severe Obesity among Adults: United States, 2017–2018." *National Center for Health Statistics Data Brief* 360.

Ho, Gloria Y. F., Hong Qian, Mimi Y. Kim, Thomas A. Melnik, Katherine L. Tucker, Ivonne Z. Jimenez-Velazquez, Robert C. Kaplan, et al. 2006. "Health Disparities between Island and Mainland Puerto Ricans." *Revista Panamericana de Salud Pública* 19 (5): 331–39.

Hummer, Robert A., Monique Biegler, Peter B. De Turk, Douglas Forbes, W. Parker Frisbie, Ying Hong, and Starling G. Pullum. 1999. "Race/Ethnicity, Nativity, and Infant Mortality in the United States." *Social Forces* 77 (3): 1083–1117.

Hummer, Robert A., Richard G. Rogers, Sarit H. Amir, Douglas Forbes, and W. Parker Frisbie. 2000. "Adult Mortality Differentials among Hispanic Subgroups and Non-Hispanic Whites." *Social Science Quarterly* 81 (1): 459–76.

Isasi, Carmen R., Guadalupe X. Ayala, Daniela Sotres-Alvarez, Hala Madanat, Frank Penedo, Catherine M. Loria, John P. Elder, et al. 2015. "Is Acculturation Related to Obesity in Hispanic/Latino Adults? Results from the Hispanic Community Health Study/Study of Latinos." *Journal of Obesity*.

James, Sherman A. 1993. "Racial and Ethnic Differences in Infant Mortality and Low Birth Weight: A Psychosocial Critique." *Annals of Epidemiology* 3 (2): 130–36.

Jasso, Guillermina, and Mark R. Rosenzweig. 1982. "Estimating the Emigration Rates of Legal Immigrants Using Administrative and Survey Data: The 1971 Cohort of Immigrants to the United States." *Demography* 19 (3): 279–90.

Jasso, Guillermina, Douglas S. Massey, Mark Rosenzweig, and James P. Smith. 2005. "The US New Immigrant Survey: Overview and Preliminary Results Based on the New-Immigrant Cohorts of 1996 and 2003." In *Immigration Research and Statistics Service Workshop on Longitudinal Surveys and Cross-Cultural Survey Design: Workshop Proceedings*, edited by Beverly Morgan and Ben Nicholson, 29–46. London: Crown Publishing.

Jerant, Anthony, Rose Arellanes, and Peter Franks. 2008. "Health Status among US Hispanics Ethnic Variation, Nativity, and Language Moderation." *Medical Care* 46 (7): 709–17.

Kaplan, Robert C., Shrikant I. Bangdiwala, Janice M. Barnhart, Sheila F. Castañeda, Marc D. Gellman, David J. Lee, Eliseo J. Pérez-Stable, Gregory A. Talavera, Marston E. Youngblood, and Aida L. Giachello. 2014. "Smoking among US Hispanic/Latino Adults: The Hispanic Community Health Study/Study of Latinos." *American Journal of Preventive Medicine* 46 (5): 496–506.

Karno, Marvin, and Robert B. Edgerton. 1969. "Perception of Mental Illness in a Mexican-American Community." *Archives of General Psychiatry* 20 (2): 233–238.

Kennedy, Steven, Michael P. Kidd, James Ted McDonald, and Nicholas Biddle. 2015. "The Healthy Immigrant Effect: Patterns and Evidence from Four Countries." *Journal of International Migration and Integration* 16 (2): 317–32.

Lopez, Mark Hugo, Jens Manuel Krogstad, and Jeffrey S. Passel. 2022. "Who Is Hispanic?" Pew Research Center, September 15. https://www.pewresearch.org/fact-tank/2021/09/23/who-is-hispanic/.

MacDorman, Marian F., T. J. Mathews. 2013. "Infant Deaths—United States, 2005-2008." *Morbidity and Mortality Weekly Report* 62 (S3): S171–75.

Markides, Kyriakos S., and Jeannine Coreil. 1986. "The Health of Hispanics in the Southwestern United States: An Epidemiologic Paradox." *Public Health Reports* 101 (3): 253–65.

Markides, Kyriakos S., and Karl Eschbach. 2005. "Aging, Migration, and Mortality: Current Status of Research on the Hispanic Paradox." *The Journals of Gerontology: Series B* 60 (S2): S68–75.

Markides, Kyriakos S., and Karl Eschbach. 2011. "Hispanic Paradox in Adult Mortality in the United States." In *International Handbook of Adult Mortality*, edited by Richard G. Rogers and Eileen M. Crimmins, 227–40. International Handbooks of Population. Dordrecht: Springer Netherlands.

Markides, Kyriakos S., and Helen P. Hazuda. 1980. "Ethnicity and Infant Mortality in Texas Counties." *Social Biology* 27 (4): 261–71.

Martell, Brandi N., Bridgette E. Garrett, and Ralph S. Caraballo. 2016. "Disparities in Adult Cigarette Smoking—United States, 2002–2005 and 2010–2013." *Morbidity and Mortality Weekly Report* 65 (30): 753–758.

Mathews, T.J., and Marian F. MacDorman. 2013. "Infant Mortality Statistics from the 2010 Period Linked Birth/Infant Death Data Set." *National Vital Statistics Reports* 62 (8): 1–26.

McDonald, James Ted, and Steven Kennedy. 2004. "Insights into the 'Healthy Immigrant Effect': Health Status and Health Service Use of Immigrants to Canada." *Social Science & Medicine* 59 (8): 1613–27.

Medina, Lauren, Shannon Sabo, and Jonathan Vespa. 2020. *Living Longer: Historical and Projected Life Expectancy in the United States, 1960 to 2060*. Suitland, MD: US Census Bureau.

National Center for Health Statistics. 1959–2020. "Multiple Cause of Death Data, National Bureau of Economic Research." CDC. https://www.nber.org/research/data/mortality-data-vital-statistics-nchs-multiple-cause-death-data (accessed on November 20, 2022).

National Center for Health Statistics. 1983–2015. "Birth Cohort Linked Birth/Infant Death Data." CDC. https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Birth_Cohort (accessed on November 20, 2022).

National Institute on Alcohol Abuse and Alcoholism. 2021. "Alcohol and the Hispanic Community." National Institute of Health. https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/

alcohol-and-hispanic-community (accessed on November 1, 2022).

Olsen, Reed, Subhasree Basu Roy, and Hui-Kuan Tseng. 2019. "The Hispanic Health Paradox for Older Americans: An Empirical Note." *International Journal of Health Economics and Management* 19 (1): 33–51.

Pablos-Méndez, Ariel. 1994. "Mortality among Hispanics." *JAMA* 271 (16): 1237–38.

Palloni, Alberto, and Elizabeth Arias. 2004. "Paradox Lost: Explaining the Hispanic Adult Mortality Advantage." *Demography* 41 (3): 385–415.

Powers, Daniel A. 2013. "Paradox Revisited: A Further Investigation of Racial/Ethnic Differences in Infant Mortality by Maternal Age." *Demography* 50 (2): 495–520.

Puyat, Joseph H. 2013. "Is the Influence of Social Support on Mental Health the Same for Immigrants and Non-immigrants?" *Journal of Immigrant and Minority Health* 15 (3): 598–605.

Riosmena, Fernando, Rebeca Wong, and Alberto Palloni. 2013. "Migration Selection, Protection, and Acculturation in Health: A Binational Perspective on Older Adults." *Demography* 50 (3): 1039–64.

Roger, Véronique L., Alan S. Go, Donald M. Lloyd-Jones, Robert J. Adams, Jarett D. Berry, Todd M. Brown, Mercedes R. Carnethon et al. 2011. "Heart Disease and Stroke Statistics—2011 Update: A Report from the American Heart Association." *Circulation* 123 (4): 18–209.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2020. "IPUMS USA: Version 10.0." IPUMS, Minneapolis, MN. https://doi.org/10.18128/D010.V12.0 (accessed on November 20, 2022).

Sorlie, Paul D., Eric Backlund, Norman J. Johnson, and Eugene Rogot. 1993. "Mortality by Hispanic Status in the United States." *JAMA* 270 (20) (1993): 2464–8.

Stephen, Elizabeth H., Karen Foote, Gerry E. Hendershot, and Charlotte A. Schoenborn. 1994. "Health of the Foreign-Born Population: United States, 1989–90." *Advance Data* 241: 1–12.

Tarraf, Wassim, William Vega, and Hector M. González. 2014. "Emergency Department Services Use among Immigrant and Non-immigrant Groups in the United States." *Journal of Immigrant and Minority Health* 16 (4): 595–606.

Vega, William A., William M. Sribney, Sergio Aguilar-Gaxiola, and Bohdan Kolody. 2004. "12-Month Prevalence of DSM-III-R Psychiatric Disorders among Mexican Americans: Nativity, Social Assimilation, and Age Determinants." *The Journal of Nervous and Mental Disease* 192 (8): 532–41.

Watson, Tara. 2014. "Inside the Refrigerator: Immigration Enforcement and Chilling Effects in Medicaid Participation." *American Economic Journal: Economic Policy* 6 (3): 313–38.

Wilson, Valerie. 2020. "Racial Disparities in Income and Poverty Remain Largely Unchanged amid Strong Income Growth in 2019." Economic Policy Institute, September 26. https://www.epi.org/blog/racial-disparities-in-income-and-poverty-remain-largely-unchanged-amid-strong-income-growth-in-2019/.

Wu, Zheng, and Christoph M. Schimmele. 2005, "The Healthy Migrant Effect on Depression: Variation over Time?" *Canadian Studies in Population* 32 (2): 271–95.

Xu, Jiaquan, Sherry L. Murphy, Kenneth D. Kochanek, and Elizabeth Arias. 2021. "Deaths: Final Data for 2019." *National Vital Statistics Reports* 70 (8).

Young, Robert P., and Raewyn J. Hopkins. 2014. "The Hispanic Paradox Further Unraveled?" *Thorax* 69 (2): 184–5.

Yue, Dahai, Petra W. Rasmussen, and Ninez A. Ponce. 2018. "Racial/Ethnic Differential Effects of Medicaid Expansion on Health Care Access." *Health Services Research* 53 (5): 3640–56.

Zhao, Fan, and Roch A. Nianogo. 2022. "Medicaid Expansion's Impact on Emergency Department Use by State and Payer." *Value in Health* 25 (4): 630–37.

Journal of Economic Perspectives—Volume 37, Number 1—Winter 2023—Pages 169–198

Hispanic Americans in the Labor Market: Patterns over Time and across Generations

Francisca M. Antman, Brian Duncan, and Stephen J. Trejo

The 2020 US Census counted 62.1 million Hispanics, representing nearly 19 percent of the US population, and almost triple the 6.5 percent share in 1980 (Jones et al. 2021; Flores, López, and Radford 2017). It has now been more than two decades since Hispanics overtook Blacks to become the largest racial-ethnic minority group in the United States (Humes, Jones, and Ramirez 2011).

The rapid growth of the Hispanic population has received a substantial amount of research attention from economists and other social scientists. To a large extent, this research focuses on the integration, experiences, and impacts of *foreign-born* Hispanics. However, two-thirds of today's Hispanics were born in the United States, and over the past two decades, Hispanic population growth has come primarily from US births rather than from immigration (Krogstad, Passel, and Noe-Bustamante 2022). We believe that US-born Hispanics have been understudied to date and that this population is ripe for future research on a wide range of topics. The ultimate impacts of Hispanic immigration on the United States depend crucially on how the US-born children, grandchildren, and later descendants of the initial arrivals fare in this country. Consider, for example, the net fiscal impacts of US immigration: how do immigrant and native-born families compare in calculations of the taxes that they

■ *Francisca M. Antman is Associate Professor of Economics, University of Colorado Boulder, Boulder, Colorado. Brian Duncan is Professor of Economics, University of Colorado Denver, Denver, Colorado. Stephen J. Trejo is Professor of Economics, University of Texas at Austin, Austin, Texas. Their email addresses are francisca.antman@colorado.edu, brian.duncan@ucdenver.edu, and trejo@austin.utexas.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.1.169.

pay minus the government benefits and services that they receive? In short-term analyses, immigrants appear to be relatively costly, but their fiscal bottom line improves dramatically in longer-term, intergenerational analyses that incorporate the future contributions made by their US-born descendants (Smith and Edmonston 1997; Blau and Mackie 2017). These specific studies included immigrants from all countries of origin, but the point is especially applicable for Hispanics, who experience unusually large advances in education and earnings between the immigrant generation and the US-born generations that follow (Cadena, Duncan, and Trejo 2015).

In this article, we offer a broad overview of the labor market performance of Hispanics, focusing primarily on men and women born in the United States. Two central questions frame our discussion. In terms of labor market skills and outcomes, what disparities persist between the descendants of Hispanic immigrants and other Americans? What are the sources of incomplete labor market integration for US-born Hispanics?

We start by highlighting critical issues that arise in the US data sources commonly used to study Hispanics; in particular, the survey responses that social science researchers use to identify Hispanics do not always match how the population describes itself. We then document how Hispanics currently compare with other Americans in terms of education, earnings, and labor supply. Next, we summarize long-term trends in the relative labor market status of Hispanics over the past half century. Finally, we consider evidence on the patterns of Hispanic progress across immigrant generations. Throughout, we emphasize important distinctions by national origin within the overall Hispanic population, draw attention to questions that are especially promising for further study, and note how future research could benefit from improvements in data collection to better understand this diverse group of people.

Identifying Hispanics in US Data Sources

The pan-ethnic label "Hispanic" refers to individuals who trace their origin or descent to Spain or to the 19 primarily Spanish-speaking countries of North, Central, and South America, along with the Caribbean.[1] Studies of labor market outcomes for US racial-ethnic groups commonly employ the large, nationally-representative

[1] Typically, the following 20 countries are considered as the origins of US Hispanics: Spain; the North American and Caribbean countries Mexico, Puerto Rico, Cuba, and the Dominican Republic; the Central American countries El Salvador, Guatemala, Honduras, Nicaragua, Panama, and Costa Rica; and the South American countries Colombia, Ecuador, Peru, Argentina, Venezuela, Chile, Bolivia, Uruguay, and Paraguay (Gratton and Gutman 2000; Rumbaut 2006). The term "Latino" is most often used as a synonym for Hispanic, although sometimes Latino is short for "Latin American," which can also include individuals with origins in relevant countries without strong ties to the Spanish language, such as Brazil. As we will discuss, in many cases the "Hispanic" or "Latino" label does not reflect how these individuals describe their own ethnic identities (Gonzalez-Barrera 2020).

microdata samples available from decennial US Censuses through the year 2000 and annual American Community Surveys (ACS) after 2000. Researchers can identify Hispanic immigrants using the information about each respondent's country of birth that has been collected starting with the 1850 Census (Humes and Hogan 2009).

However, it is much less straightforward to identify US-born individuals of Hispanic *ancestry* consistently and comprehensively through time. For Censuses before 1970, various pieces of available information (for example, countries of birth of the respondent and the respondent's parents, mother tongue, and having a Spanish surname) can imperfectly identify subsets of the Hispanic population (Gratton and Gutmann 2000). Imputed measures of Hispanic origin, based on methods developed by Gratton and Gutmann (2000), are available through the Integrated Public Use Microdata Series (IPUMS) (Ruggles et al. 2022). While imperfect, these measures have facilitated research on the Hispanic population in earlier waves of the US Census (for example, Antman and Cortes forthcoming).

The 1970 Census introduced the first question that allowed respondents to self-identify as Hispanic. Because this question was added at the last minute in response to political pressure and could not be pretested, it suffered from data quality problems that might have been avoided. For example, the wording of the Hispanic origin question led many non-Hispanic respondents living in the central and southern regions of the United States to mistakenly report that they were members of the Hispanic category "Central or South American" (Siegel and Passel 1979; Humes and Hogan 2009).

Starting with the 1980 Census, the Hispanic origin question adopted the improved format that it has maintained, with relatively minor changes, through later years. The 2019 American Community Survey, for example, first asks whether the respondent is "of Hispanic, Latino, or Spanish origin." If the answer is "yes," then a box is checked to indicate whether the specific Hispanic national origin group is Mexican or Mexican American, Puerto Rican, Cuban, or "another Hispanic, Latino, or Spanish origin." People who check this last box are instructed to "print origin, for example, Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on." Space is provided for this write-in response. Many other surveys and most contemporary empirical analyses of US-born Hispanics make use of this particular Hispanic origin question, or something very similar, to identify persons of Hispanic ancestry.

Hispanic Origin versus Race

It is important to emphasize that this question about Hispanic origin is separate from the Census question about race that is typically used to identify other minority groups, such as Blacks, Asians, and American Indians. The Hispanic origin question and the race question collect independent information about racial and ethnic identity in the Census, American Community Survey, and many other surveys. In the Census and ACS, respondents are explicitly instructed to answer "BOTH" the question about Hispanic origin and the question about race, and they are told that

"for this survey, Hispanic origins are not races." According to federal standards for the collection of data on race and ethnicity, the Hispanic origin question provides information regarding ethnicity, which these standards consider to be a distinct concept from race (Humes, Jones, and Ramirez 2011). In line with this view, the detailed ACS instructions to respondents state that Hispanics "may be of any race."

However, many individuals see race and ethnicity as inextricably linked, and Hispanic respondents in particular may not see themselves reflected in the official response options available for the race question, which include White, Black or African American, American Indian or Alaska Native, several Asian categories, and "some other race" (where individuals can write in their own response) (Jones et al. 2021). For example, the definition of the American Indian race category currently includes "original peoples of North and South America (including Central America)," but also suggests that individuals selecting that category maintain "tribal affiliation or community attachment" (Office of Management and Budget 1997; see Antman and Duncan 2021 for further discussion). Complicating matters further, the race and Hispanic origin questions are distinct from a separate question about ancestry. Antman (2022) provides further discussion of the complexities of racial and ethnic self-identification for the Hispanic population and the implications for research and policy.

In the 2010 Census, for example, 53 percent of Hispanics reported a race of "White," but another 37 percent chose a race response deemed invalid by the Census Bureau. Typically, these individuals wrote in a Hispanic national origin group (like Mexican or Dominican) or a pan-ethnic Hispanic label (like Hispanic or Latino). In editing the data, the Census Bureau classifies such individuals as "some other race;" as a result, Hispanics made up 97 percent of the "some other race" population in the 2010 Census (Humes, Jones, and Ramirez 2011). Similar analysis of the 2020 Census indicates that "some other race" (again largely consisting of Hispanic individuals) is now the second-largest US racial group after White, raising calls for the race question to be modified (Wang 2021; Bahrampour 2021).

There are other consequential differences between the race and Hispanic origin questions in the Census and the American Community Survey. For example, the race question has been asked since the first decennial Census in 1790, and therefore data for Blacks can be tracked for more than two centuries. In contrast, as noted above, the Hispanic origin question is barely 50 years old. In addition, beginning with the 2000 Census, the race question permits multiple responses—respondents are instructed to "mark one or more races" (Humes and Hogan 2009). The Hispanic origin question, however, continues to solicit and report only a single response for Hispanic national origins; for example, respondents cannot identify as having both Puerto Rican and Cuban ancestry.

Identifying Immigrant Generation

Because many US-born Hispanics are only a generation or two removed from their ancestors who first moved to this country, it is useful for researchers to distinguish Hispanics by their immigrant generation, where the first generation consists of the foreign-born immigrants themselves, the second generation consists of US-born

individuals with at least one foreign-born parent, the third generation are US-born individuals with US-born parents and at least one foreign-born grandparent, and so on. To begin to make these distinctions requires information, at the very least, on the countries of birth of respondents and their parents. Such information is available in Census data from 1880 to 1970, but starting in 1980 the Census stopped asking about the countries of birth of respondents' parents. As a result, in contemporary data from the decennial Census or the annual American Community Survey, adults can be distinguished by nativity—that is, foreign-born versus US-born—but it is not possible to make further distinctions with respect to immigrant generation among US-born adults.

In 1994, the Current Population Survey (CPS) began collecting information on the countries of birth of respondents and their parents. As a result, this survey currently provides the best large-scale, nationally representative US data for investigating how outcomes vary by immigrant generation. In contemporary data from the CPS, Hispanic adults can be assigned to the first, second, and "third-plus" (that is, third and higher) generations. First-generation Hispanics were born in a Hispanic country. Second-generation Hispanics are US-born individuals with a parent born in a Hispanic country. Third-plus generation Hispanics are those who were born in the United States, have two US-born parents, and self-identify as Hispanic in response to the Hispanic origin question. Notice that first- and second-generation Hispanics are identified by a relatively objective question about birthplaces, whereas third-plus-generation Hispanics must subjectively identify as being Hispanic. We return to this point below when discussing how to interpret observed patterns across immigrant generations in these data.

Current Educational Attainment and Labor Market Status of Hispanics

We use data from the 2019 American Community Survey to describe the relative labor market status of Hispanics just prior to the onset of the COVID-19 pandemic.[2] We begin with educational attainment, because schooling is a fundamental determinant of economic success, social status, health, family stability, and life opportunities (Oreopoulos and Salvanes 2011; Hout 2012). In addition, information on education is available for all adults, whereas earnings data are available only for those currently working.

Education

Among individuals ages 25–59 (including those residing in group quarters), Table 1 reports differentials, relative to US-born non-Hispanic Whites, in average

[2] All of the Census and American Community Survey microdata used in this article are from IPUMS-USA (Ruggles et al. 2022). The Current Population Survey microdata used later in the paper are from IPUMS-CPS (Flood et al. 2022).

Table 1

Years of Schooling Differentials, Relative to US-Born Non-Hispanic Whites, by Race-Ethnicity and Nativity, 2019

	Nativity	
Race-Ethnicity	*Foreign-Born*	*US-Born*
Non-Hispanic Blacks		–0.85
		(0.01)
All Hispanics	–3.16	–1.11
	(0.02)	(0.01)
Hispanics by National Origin:		
Mexican	–4.02	–1.28
	(0.02)	(0.01)
Puerto Rican	–1.33	–1.09
	(0.04)	(0.03)
Cuban	–0.98	0.06
	(0.05)	(0.06)
Central American	–4.23	–0.89
	(0.05)	(0.06)
South American	–0.29	0.25
	(0.04)	(0.06)
Dominican	–1.91	–0.61
	(0.06)	(0.08)
Other Hispanic	–1.89	–1.00
	(0.11)	(0.04)

Source: 2019 American Community Survey microdata from IPUMS-USA.
Notes: For selected racial-ethnic/nativity groups, this table reports differentials (relative to US-born non-Hispanic Whites) in average years of schooling. Standard errors are shown in parentheses. The differentials are estimated coefficients on indicators for the relevant racial-ethnic/nativity groups from a least squares schooling regression that also includes age fixed effects. Sampling weights are employed. The regression sample includes men and women ages 25–59 who are members of one of the following racial-ethnic/nativity groups: Hispanics (foreign-born and US-born), US-born non-Hispanic Blacks, and US-born non-Hispanic Whites.

years of schooling for selected groups defined by race-ethnicity and nativity.[3] These calculations pool together men and women, but patterns are similar when the data are disaggregated by sex. The reported differentials are estimated by a least squares regression that controls for age and employs sampling weights. Among the US-born, Hispanics skew younger than other groups, with an average age in our sample of 38.0 compared with 42.4 for non-Hispanic Whites and 40.7 for non-Hispanic

[3]Hispanics and Hispanic national origin groups are identified from the Hispanic origin question in the American Community Survey, and non-Hispanic Blacks and non-Hispanic Whites are identified from the race question. Foreign-born individuals are those born outside of the United States. Here, we include as foreign-born those born in Puerto Rico and other US territories and outlying areas. The Hispanic national origin groups listed separately in Table 1 are those with the largest US populations; the residual category "Other Hispanic" captures only 5 percent of Hispanics.

Blacks. Controlling for age in the schooling regression accounts for general trends in educational attainment across birth cohorts, and in particular it adjusts for the fact that older individuals grew up when people typically acquired less schooling than they have in recent years. Sampling weights help improve the precision and national representativeness of the estimates.

Hispanics as a group possess relatively low levels of educational attainment. Compared to US-born non-Hispanic Whites, Hispanic immigrants have a schooling deficit of over three years, and the corresponding deficit for US-born Hispanics is more than a year. Even among the US-born, average years of schooling of Hispanics trails that of Blacks by a quarter of a year. These aggregate statistics for Hispanics, however, conceal enormous diversity across national origin groups, particularly for the foreign-born. Mexican and Central American immigrants have schooling deficits exceeding four years, more than quadruple the corresponding gaps for foreign-born Cubans and South Americans. Among the US-born, Cubans and South Americans attain educational parity with non-Hispanic Whites, whereas other Hispanic national origin groups exhibit deficits clustered around one year. For schooling as well as for other outcomes, it is important to keep in mind that, because Mexican-origin individuals constitute half of Hispanic immigrants and two-thirds of US-born Hispanics (in the samples in Table 1), overall patterns for Hispanics will largely reflect those of this dominant national origin group.

Earnings

Table 2 presents analogous differentials in annual earnings, separately for men and women, where annual earnings include wage and salary income and (nonnegative) self-employment income received over the previous twelve months. Here, the samples are restricted to individuals with positive earnings. The regressions that estimate these differentials use the natural logarithm of earnings as the dependent variable and include control variables for state of residence as well as for age. Hispanics are clustered geographically, with three-fifths of this population located in just four states in 2019: California (25.7 percent), Texas (19.2 percent), Florida (9.4 percent), and New York (6.2 percent) (Krogstad 2020). Controlling for state of residence in the earnings regressions helps to account for geographic variation in the cost-of-living and local economic conditions.

Earnings differentials potentially provide a more complete measure of labor market disparities than do educational differences, because earnings represent the market's valuation of a worker's entire package of abilities and attributes, including those for which data are often lacking (like family background or the quality of schooling). In addition, earnings differentials may capture discrimination and the unequal treatment of particular groups in the labor market. At the annual level, earnings differentials also reflect differences between groups in the average number of hours worked per year.

Because the regressions estimating earning differentials employ the log of earnings as the dependent variable, the estimated differentials represent log point differences, which closely approximate percentage differences for differences

Table 2

Log Annual Earnings Differentials, Relative to US-Born Non-Hispanic Whites, by Race-Ethnicity, Nativity, and Sex, 2019

	Men, by Nativity		*Women, by Nativity*	
Race-Ethnicity	*Foreign-Born*	*US-Born*	*Foreign-Born*	*US-Born*
Non-Hispanic Blacks		–0.471		–0.171
		(0.007)		(0.006)
All Hispanics	–0.495	–0.296	–0.520	–0.189
	(0.005)	(0.006)	(0.007)	(0.007)
Hispanics by National Origin:				
Mexican	–0.539	–0.325	–0.594	–0.229
	(0.007)	(0.008)	(0.010)	(0.009)
Puerto Rican	–0.394	–0.319	–0.359	–0.204
	(0.025)	(0.017)	(0.026)	(0.018)
Cuban	–0.357	–0.053	–0.303	0.119
	(0.021)	(0.042)	(0.024)	(0.029)
Central American	–0.566	–0.249	–0.610	–0.126
	(0.011)	(0.025)	(0.017)	(0.027)
South American	–0.278	–0.062	–0.327	0.047
	(0.016)	(0.029)	(0.017)	(0.029)
Dominican	–0.504	–0.356	–0.555	–0.208
	(0.026)	(0.044)	(0.027)	(0.043)
Other Hispanic	–0.453	–0.287	–0.375	–0.176
	(0.042)	(0.023)	(0.048)	(0.026)

Source: 2019 American Community Survey microdata from IPUMS USA.
Notes: For selected racial-ethnic/nativity groups, this table reports differentials (relative to US-born non-Hispanic Whites) in the natural logarithm of average annual earnings, separately for men and women. Standard errors are shown in parentheses. The differentials are estimated coefficients on indicators for the relevant racial-ethnic/nativity groups from least squares log earnings regressions, run separately by sex, that also include fixed effects for age and state of residence. Sampling weights are employed. The regression samples include individuals ages 25–59 with positive earnings who are members of one of the following racial-ethnic/nativity groups: Hispanics (foreign-born and US-born), US-born non-Hispanic Blacks, and US-born non-Hispanic Whites.

on the order of 20 log points or less in absolute value. For larger differences, the implied percentage difference can be calculated as $[e^{(.01)x} - 1] \times 100$, where x is the log point difference. Foreign-born Hispanic men, for example, have an earnings deficit of 50 log points (39 percent) relative to US-born non-Hispanic White men, which is similar to the corresponding deficit of 52 log points (41 percent) for foreign-born Hispanic women relative to non-Hispanic White women. These earnings deficits are considerably smaller but still substantial for US-born Hispanics: 30 log points (26 percent) for men and 19 log points (17 percent) for women. By comparison, the earnings deficit for US-born Black men is 47 log points (37 percent). Among US-born men, therefore, Hispanics earn on average about one-fourth less than non-Hispanic Whites of the same age who live in the same state—a sizeable earnings disadvantage, but markedly smaller than the comparable

deficit of 37 percent for Black men. Among US-born women, by contrast, earnings gaps relative to non-Hispanic Whites are similar for Hispanics (19 log points) and Blacks (17 log points).

Employed Hispanic men and women work about the same number of hours per year as their non-Hispanic White counterparts, so differentials in average weekly or hourly earnings are similar to those reported in Table 2 for annual earnings. The same is true for Black women, but for employed Black men, relatively fewer hours of work account for over one-quarter of their annual earnings deficit.

A fundamental source of the earnings deficits shown in Table 2 is the education gaps reported in Table 1. Among US-born Mexican Americans, for example, controlling for years of schooling shrinks these annual earnings differentials by 57 percent (see panel B of Figure 1 below). Education also plays a key role in explaining lower earnings for Blacks, but for Black men in particular this role is notably smaller, with their schooling gap accounting for only about one-fifth of their earnings deficit. Several studies have noted the critical importance of shortfalls in education and other observable measures of skill in explaining the earnings disadvantage of Hispanics in general (Smith 2001; Duncan, Hotz, and Trejo 2006) and Mexican Americans in particular (Trejo 1997; Antecol and Bedard 2002).

Lessons for Research on Hispanics

Together, Tables 1 and 2 highlight several essential points about the current labor market standing of Hispanics. An obvious point, but one often neglected in media reports and sometimes even in academic analyses, is the crucial distinction between foreign-born and US-born Hispanics. Many Hispanic immigrants arrive in the United States with rudimentary levels of schooling and little or no knowledge of the English language, so it is not surprising that they usually earn substantially less than other Americans. Nonetheless, despite their low skills, Hispanic immigrants maintain high rates of paid employment (Cadena, Duncan, and Trejo 2015). Compared to their immigrant ancestors, US-born Hispanics experience large improvements in educational attainment and earnings. Analyses that pool together foreign-born and US-born Hispanics create misleading portraits of Hispanic labor market status.

Another important point is the advisability, whenever possible, of disaggregating Hispanics by national origin. While Hispanic national origin groups share the same ancestral language and perhaps some cultural traditions, they differ widely in terms of the source country environments they left behind, their migration histories and context of arrival (for example, as political refugees, undocumented immigrants, or legal admissions), where they concentrate geographically within the United States, skin color and racial identity, and myriad other factors (Rumbaut 2006). In light of these fundamental differences, the diversity of educational and earnings outcomes across national origin groups reported in Tables 1 and 2 is perhaps unsurprising.

Finally, Tables 1 and 2 remind us of the substantial labor market disparities faced by Hispanics today, even among those who were born and raised in the United States.[4] Substantial educational deficits exist for US-born Hispanics in most national origin groups, and these educational deficits account for much of the corresponding earnings gaps. Consequently, there could be large benefits from research that improves our understanding of why schooling remains persistently low for US-born Hispanics and what interventions could raise their educational attainment. It would also be valuable to increase our understanding of other sources of Hispanic earnings gaps besides schooling deficits. Although one of the earliest audit studies of hiring discrimination focused on young Hispanic men (Kenney and Wissoker 1994), there has been relatively little work since then using research designs that can credibly distinguish discrimination from other potential sources of labor market disparities for Hispanics. An influential study by Bertrand and Mullainathan (2004) estimates hiring discrimination against African Americans by responding to help-wanted ads with fictitious resumes, where the names on the resumes have been randomly assigned to signal the race of the applicant. This approach could be adapted to learn more about hiring discrimination against Hispanics.

Trends over the Past Half Century

Numerous studies have tracked long-term labor market trends for African Americans going back to before World War II (for example, Smith and Welch 1989; Collins and Margo 2006; Neal 2006; Bayer and Charles 2018), but—with the notable exception of Smith (2003; 2006)—there is a dearth of comparable empirical research for Hispanics. In part, this may reflect data limitations discussed earlier: the Hispanic origin question that can identify US-born Hispanics was not introduced until the 1970 Census and did not become widely used until the 1980 Census. In this section, we summarize trends in the relative educational attainment and earnings of Hispanics over the past 50 years.

We will focus on US-born Mexican Americans, rather than on Hispanics more broadly defined. Given the sizeable differences in education and earnings across national origin groups that were documented in the previous section, overall trends for Hispanics confound shifts over time in the national origin composition of this population with the changes taking place within specific national origin groups. Restricting the sample to Mexican Americans, the largest Hispanic national origin group, avoids this difficulty. Also, the ambiguities discussed earlier with the Hispanic origin question in the 1970 Census are less problematic for Mexican Americans.

[4]More recent data tracking labor market outcomes during the COVID-19 pandemic suggest that these disparities were exacerbated by the pandemic and associated lockdowns, which hit Hispanic households especially hard due to a variety of factors including their occupations and industries of employment, as well as their preexisting socioeconomic vulnerabilities (Gould, Perez, and Wilson 2020).

Nonetheless, the basic trends for all Hispanics turn out to be similar to those reported here for Mexican Americans.

Education Trends

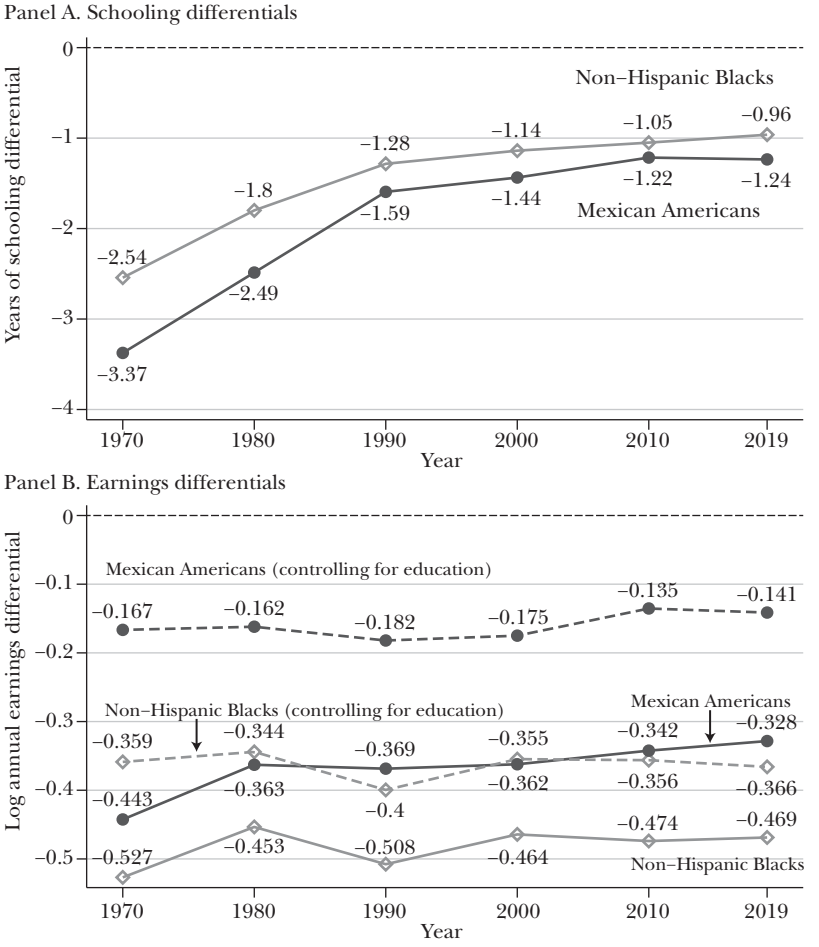
For US-born men ages 25–59, Figure 1 reports schooling differentials (panel A) and log annual earnings differentials (panel B) for Mexican Americans and Blacks (relative to non-Hispanic Whites) at decade intervals from 1970 through 2019. The schooling differentials control for age and the earnings differentials also control for state of residence and sometimes for years of education. To minimize potential biases in the earnings differentials arising from selective labor force participation, this figure only reports calculations for men. The education and earnings trends for Mexican-American women, however, are similar to those shown for Mexican-American men.

The top panel of Figure 1 documents large educational gains for US-born Mexican-American and Black men during the past half century. Over this period, Mexican Americans closed their schooling gap relative to non-Hispanic White men by 2.1 years, while the corresponding gain for Blacks was 1.6 years. For each group, these gains erase more than 60 percent of their initial educational deficit. For each group, about 80 percent of the total gains were realized within the initial 20-year period from 1970 to 1990, with a much slower rate of educational convergence evident over the most recent 30 years.

These educational gains for Mexican Americans and Blacks are driven by relative improvements in their rates of high school completion. In 1970, the age-adjusted rate of high school completion for Mexican-American men was 34 percentage points below that of non-Hispanic White men, but this deficit shrank steadily across decades down to 10 percentage points in 2019. For Black men, the corresponding disparity in high school completion declined from 29 percentage points in 1970 to 6 percentage points in 2019. Examining data by birth cohort and not counting GED recipients as high school graduates, Murnane (2013) also reports evidence of relative improvement in high school graduation for Hispanics and Blacks.

At the upper end of the educational distribution, however, college completion rates have diverged rather than converged. All racial-ethnic groups have experienced rapid growth over the past 50 years in the share of adults completing a bachelor's degree, but this growth has been more rapid for non-Hispanic Whites than for Mexican Americans and Blacks. Among US-born men, for example, the Mexican-American deficit in college completion (relative to non-Hispanic Whites) widened from 15 percentage points in 1970 to 20 percentage points in 2019. For Black men, this deficit increased from 13 to 18 percentage points over the same five decades. Finding a way to reverse these patterns is essential if Hispanics and Blacks are to continue closing their educational gaps relative to other Americans. Bleemer (2022) provides evidence that, prior to their elimination in 1998, affirmative action policies at the University of California system benefited Hispanic students by increasing their access to better-quality colleges and thereby raising graduation rates and boosting future earnings. The future of affirmative action policies in

Figure 1
Schooling and Earnings Differentials, Relative to US-Born Non-Hispanic Whites, for US-Born Mexican-American and Black Men, 1970–2019



Source: 1970–2000 Census and 2010 and 2019 American Community Survey microdata from IPUMS-USA.
Notes: For US-born men who are Mexican American or non-Hispanic Black, this figure shows differentials (relative to US-born non-Hispanic Whites) in average years of schooling (panel A) and the natural logarithm of average annual earnings (panel B). The differentials are estimated coefficients on indicators for the relevant racial-ethnic groups from least squares schooling (panel A) or log earnings (panel B) regressions, run separately for each survey year. These regressions include fixed effects for age (in both panels) and state of residence (in panel B). The dashed plot lines in panel B are from earnings regressions that also control for years of education. Sampling weights are employed. The samples for the schooling regressions include US-born men ages 25–59 who are members of one of the following racial-ethnic groups: Mexican Americans, non-Hispanic Blacks, and non-Hispanic Whites. The samples for the earnings regressions are further restricted to individuals with positive earnings.

college admissions, currently being debated before the US Supreme Court, is thus highly relevant to the evolution of these disparities. These issues have become more urgent due to the COVID-19 pandemic, which has reduced college enrollment rates for all racial and demographic groups (Sedmak 2021), but perhaps especially so for Hispanic students (Ahn and Dominguez-Villegas 2022).

For Blacks, a substantial research literature has described their long-term educational progress and the sources of this progress (Collins and Margo 2006; Neal 2006). This work has documented gains in both the quantity and quality of schooling received by Blacks (Smith and Welch 1989; Card and Krueger 1992), and it has also assessed the impacts of specific interventions such as court-ordered school desegregation (Guryan 2004; Reber 2010; Lutz 2011) and philanthropic efforts to improve school access and resources (Aaronson and Mazumder 2011). In contrast, relatively little research explores these issues for Hispanics.

Again, this lack of research may stem in part from data challenges which make it difficult to identify the Hispanic population in pre-1970 Censuses. In addition, there is a lack of official documentation surrounding the educational treatment of Hispanic children in US schools. Historical accounts document systemic discrimination and segregation of Hispanic children throughout many counties in the US Southwest in the early twentieth century (Rangel and Alcala 1972; Wollenberg 1976), but official records are notably absent. Despite these challenges, Antman and Cortes (forthcoming) estimate the long-run impacts on Hispanic educational attainment of the court-ordered end to de jure segregation of Hispanic school children in California which came about with the 1947 *Mendez v. Westminster* (161 F.2d 774 [9th Cir. 1947]) decision. Antman and Cortes find that desegregation resulted in about 0.9 additional years of schooling—an almost 9 percent increase relative to educational attainment for Hispanic cohorts who started school in the segregated era in counties where segregation was most likely to have occurred. The parallels with the 1954 *Brown v. Board of Education of Topeka* (347 U.S. 483 [1954]) decision applying to African American school segregation are significant, although there is a relative abundance of official records documenting segregation for Blacks in the US South which facilitates research on this particular group. Much remains to be learned about the factors underlying Hispanic educational progress.

Earnings Trends

Panel B of Figure 1 presents log annual earnings differentials for US-born Mexican-American and Black men (relative to non-Hispanic White men) over the past five decades. Two varieties of earnings differentials are displayed. The solid plot lines show earnings differentials that condition on age and state of residence, whereas the dashed plot lines show earnings differentials that also control for years of education.

First, consider the earnings differentials that do not condition on education. In 1970, the earnings of Mexican-American workers trailed the earnings of non-Hispanic White workers by 44 log points, but this gap had closed to 36 log points by 1980. There was little change in the earnings deficit over the next two

decades (1980–2000) and then a gradual, modest decline from 36 to 33 log points during the most recent period (2000–2019). At each point in time, the earnings deficits for Black men exceeded those for Mexican-American men by at least nine log points. After benefitting from sizeable reductions in their earnings deficits between 1970 and 1980, both groups experienced little or no earnings convergence with non-Hispanic White men over the subsequent 40 years.

In Figure 1, why did the educational convergence shown in panel A fail to produce more earnings convergence in panel B? The US wage structure has changed dramatically since 1980, with large increases in overall earnings inequality and sharply rising returns to labor market skills, including education (Autor, Katz, and Kearney 2008). As Mexican-American and Black men were reducing their educational deficits relative to non-Hispanic White men, the labor market penalty associated with a given deficit was increasing, and these offsetting forces combined to produce little net change in the corresponding earnings deficits. In effect, Mexican-American and Black workers have been swimming upstream against these ongoing changes in the US economy. Bayer and Charles (2018) provide a detailed analysis of how these and other forces have affected the relative earnings of Black men since 1940. Comparable analyses for Mexican Americans or other Hispanic groups have yet to be done.

Let us now turn to the earnings differentials in panel B of Figure 1 that condition on years of education—the dashed plot lines. Over the past 50 years, these schooling-adjusted earnings deficits (relative to non-Hispanic Whites) have hovered within a relatively narrow range for both Mexican Americans (14–18 log points) and Blacks (34–40 log points), resulting in deficits that are always less than half as large for Mexican-American men as for Black men. In addition, comparing the dashed lines that condition on schooling with the solid lines that do not, we see that controlling for education reduces earnings deficits to a much greater extent for Mexican Americans than for Blacks, as was mentioned earlier. In summary, for both Mexican-American and Black men, earnings disparities relative to non-Hispanic White men have persisted with only modest declines over the past five decades. These earnings deficits are substantially smaller for Mexican Americans than for Blacks, even more so after accounting for the corresponding schooling deficits. Nonetheless, the persistence of sizeable, unexplained earnings deficits for Mexican Americans and other Hispanic national origin groups deserves further attention.

Another question that merits further study is the increasing labor force activity of US-born Mexican-American women. Among all native-born women ages 25–59, for example, in 1980 the percent of Mexican Americans who had worked in the previous year was 8 percentage points lower than that of non-Hispanic Whites, after controlling for age and state of residence. This differential shrank to 3 percentage points by 2019, with most of the decline occurring after 2000. Moreover, within the sample of women who worked in the previous year, similar calculations reveal that in 1980 Mexican Americans averaged 5 percent fewer annual hours of work than non-Hispanic Whites, but by 2019 they averaged 2 percent *more* annual hours of work than non-Hispanic Whites. Over this period, Mexican-American women

experienced even larger employment and work hours gains relative to Black women. Therefore, among US-born women, Mexican Americans have increased their labor force activity on both the extensive and intensive margins, and remaining differences in employment and work hours across racial-ethnic groups are small compared to 30 or 40 years ago. More generally, distinctions within the Hispanic population are again critical for research in this area, because labor supply varies considerably between US- and foreign-born Hispanics, as well as between documented and undocumented immigrants. For example, Borjas (2017) finds higher labor force participation rates for undocumented immigrant men and lower labor force participation rates for undocumented immigrant women, relative to the non-Hispanic White population.

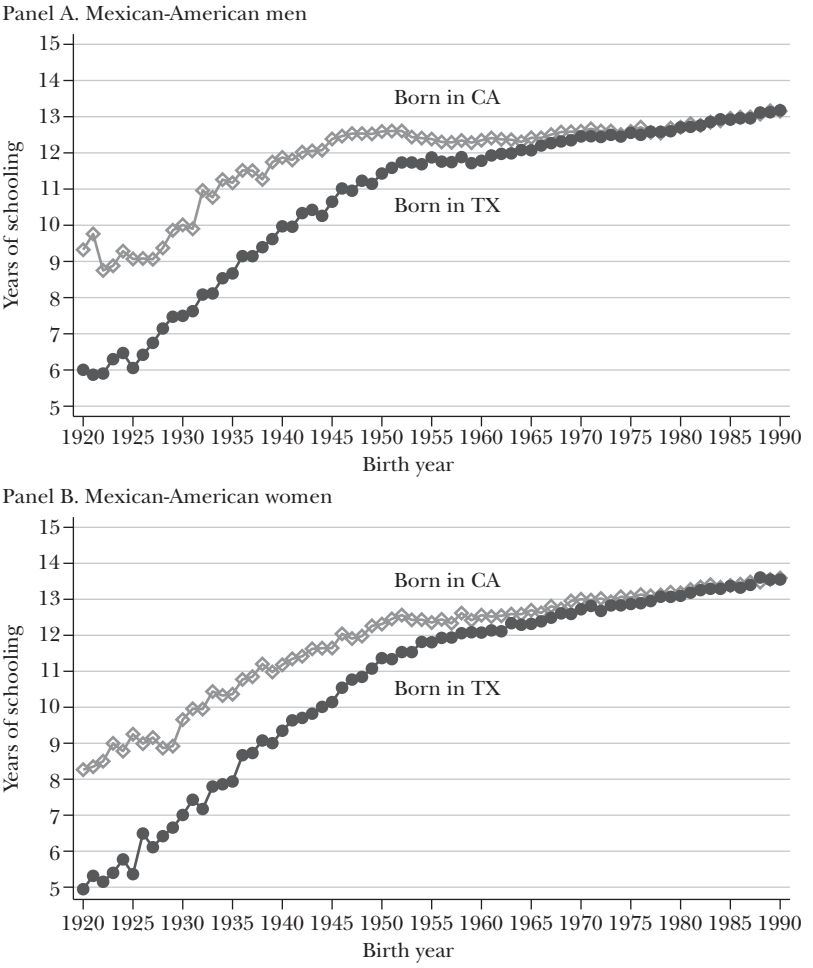
California and Texas

For Mexican Americans, California and Texas are important regional focal points, with over two-thirds of the US-born population originating in just these two states. Figure 2 highlights historical differences in educational attainment between Mexican Americans born in California and those born in Texas. For these calculations, we pool together microdata from the American Community Survey for 2006–2019, as well as from the decennial Censuses of 1970–2000, and we estimate how average completed years of schooling has evolved across cohorts defined by birth year.

Panel A of Figure 2 displays the resulting estimates for men, and panel B shows the estimates for women. For most of the twentieth century, schooling levels were much lower for Mexican Americans born in Texas than for those born in California. Among Mexican Americans born in the early 1920s, Californians average three more years of schooling than do Texans. This enormous initial gap steadily narrows over time, however, until it disappears for those born in the mid-1970s and later. Historically, schooling levels for non-Mexicans were also somewhat lower for those born in Texas; nonetheless, even when comparing Mexican Americans to others within the same state of birth, the resulting educational deficits of Mexican Americans relative to non-Hispanic Whites (or relative to Blacks) were much larger for Texans than for Californians. Kosack and Ward (2020) report a similar finding with respect to occupational standing in 1940. Relative to non-Mexican Whites, Mexicans Americans were particularly disadvantaged in Texas compared with California.

What were the sources of the large initial educational disadvantage for Mexican Americans born in Texas, and why did this disadvantage fade? These questions are ripe for further study, perhaps by taking advantage of the full-count data that have become available for the 1940 and earlier US Censuses (Ruggles et al. 2021). As noted earlier, unlike contemporary Census data that only report country of birth for the respondent, Censuses up through 1970 also report the countries of birth of each respondent's parents. This information enables identification of second-generation Mexican-American adults (US-born individuals with a parent born in Mexico). Merging the information of parents and their children residing in

Figure 2
Average Years of Schooling of US-Born Mexican Americans, by Birth Year, Sex, and Birth State



Source: 1970–2000 Census and 2006–2019 American Community Survey microdata from IPUMS-USA.
Notes: For Mexican Americans born in California and in Texas, this figure shows predicted average years of schooling at age 35 for each birth year, separately for men and women. In each survey, we limit our analysis sample to US-born individuals between the ages of 25–59, including those residing in group quarters. Because we know each respondent's current age but not their birthday, we approximate the birth year of each individual as Birth Year = Survey Year – Age – 1. We limit this analysis to individuals with birth years between 1920 and 1990, which are the birth years we observe in at least two different survey years. For any given birth cohort, average educational attainment tends to rise with the age at which the cohort is observed. To standardize for age effects, we report predicted years of schooling at age 35 for each birth year. These predictions derive from least squares regressions of schooling on a quartic in age and dummy variables identifying the relevant cells defined by sex, state of birth, and year of birth.

the same household enables identification of third-generation Mexican-American children (US-born individuals with US-born parents and a grandparent born in Mexico).

Previous research has noted the historical schooling disadvantage of Texas-born Mexican Americans (Grebler, Moore, and Guzman 1970; Bean, Brown, and Bachmeier 2015), and historians have documented that, until the implementation of civil rights reforms, Mexican Americans in Texas often attended schools that were more segregated and given fewer resources than schools that Mexican Americans attended in other states (Rangel and Alcala 1972; Montejano 1987). Preliminary analyses with 1930 and 1940 Census data suggest a related factor that might be important. School enrollment rates were much lower for Mexican-American children born in Texas than for those born in California, and enrollment rates were especially low for Texas-born Mexican Americans residing on farms. Historical accounts highlight the particularly severe obstacles to schooling faced by Mexican-American children in Texas farm areas (Montejano 1987). In 1940, about a quarter of Texas-born Mexican-American children lived on farms, compared with only 10 percent of their California-born counterparts. Therefore, the movement away from farms (and farm schools) that accelerated after World War II may have produced especially large educational gains for Texas-born Mexican Americans. More generally, disproportionately large shares of Mexican Americans—US-born individuals as well as immigrants—worked in agriculture and lived in farm communities during the first half of the 1900s (Gratton and Merchant 2015). As a result, the dramatic decline in agricultural employment and the transformation of rural communities that took place throughout the century potentially impacted Mexican Americans in a variety of significant ways. These topics deserve further study.

Generational Progress

Most Hispanics are no more than a generation or two removed from their family's migration to the United States. For example, based on data from the Current Population Survey described in more detail below, over half of Hispanics ages 18–59 are first generation (that is, foreign-born immigrants) and another 23 percent are second generation (that is, the US-born child of an immigrant). Only 23 percent of Hispanics are third generation or higher (that is, US-born individuals with US-born parents, so that the initial family member immigrating to the United States was a grandparent or earlier ancestor). By contrast, 91 percent of non-Hispanic Whites are third generation or higher. As discussed earlier, US-born Hispanics exhibit substantial improvements in schooling and earnings relative to Hispanic immigrants, but large socioeconomic disparities remain between US-born Hispanics and non-Hispanic Whites. Given the magnitude of the initial human capital deficits possessed by most Hispanic immigrants, it may take their US-born descendants more than one generation to catch up. Does

continued progress across immigrant generations become evident when US-born Hispanics are distinguished by how many generations their family has lived in this country?

Education and Immigrant Generation

To shed light on this question, we turn to monthly data from the Current Population Survey (CPS) from 2003 to 2019.[5] As mentioned previously, a key feature of this data source is that, starting in 1994, it includes information about parental countries of birth that is currently missing from the Census and the American Community Survey. As a result, adult respondents in CPS data can be assigned to three categories of immigrant generation: first, second, and third-plus generation (that is, third or higher generation). These generation categories can be further disaggregated by ethnicity and race, with Hispanics and Hispanic national origin groups identified using the Hispanic origin question, and non-Hispanic Blacks and non-Hispanic Whites identified from the race question.

Table 3 shows how average years of schooling varies with immigrant generation for Hispanics overall and also for the Mexican national origin group. For comparison purposes, Table 3 also reports average schooling for non-Hispanic White and non-Hispanic Blacks in the third-plus generation. Panel A presents these statistics separately for men and women ages 25–59. Because the national origin composition of the Hispanic population differs significantly across generations, we focus on the results for the Mexican national origin group. In general, however, the patterns are similar for Hispanics overall.

For both men and women, panel A of Table 3 indicates that educational attainment is more than three years higher for second-generation Mexican Americans than for their first-generation counterparts. Progress seems to stall after the second generation, however, with no further improvement for the third-plus generation. Third-plus-generation Mexican Americans have schooling deficits relative to non-Hispanic Whites of more than a year, and they also maintain significant deficits relative to Blacks (of more than a quarter of a year for men and almost half a year for women). Analyses of how earnings vary with immigrant generation for Hispanics overall, and for Mexicans in particular, reveal a pattern similar to what panel A of Table 3 shows for education: much higher earnings for the second generation compared with the first, but minimal additional gains beyond the second generation,

[5] Specifically, we use outgoing rotation group microdata from the Current Population Survey (CPS). We start the sample in 2003 to coincide with the introduction of a revised CPS questionnaire that improved the quality of the data collected on Hispanic origin (see the appendix in Duncan and Trejo 2016). To avoid duplicate observations on a given individual, we use only data from the first time a household appears in the outgoing rotation group samples (that is, only data from the fourth month that a household appears in the CPS sample). The sampling universe for the CPS is the civilian noninstitutionalized population of the United States, which can create biases for groups with relatively high rates of incarceration, such as young Black males (Pettit 2012). The data from the Census and American Community Survey used in previous sections include people living in group quarters such as prisons, and therefore these data mitigate such biases.

Table 3

Average Years of Schooling, by Race-Ethnicity, Age, Sex, and Immigrant Generation

	Men, by Immigrant Generation			*Women, by Immigrant Generation*		
Race-Ethnicity and Age	*First*	*Second*	*Third+*	*First*	*Second*	*Third+*
Panel A. Ages 25–59						
Hispanic	10.38	13.00	12.85	10.78	13.26	13.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Mexican	9.61	12.73	12.71	9.80	12.92	12.87
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Non-Hispanic White			13.84			14.08
			(0.003)			(0.003)
Non-Hispanic Black			13.00			13.36
			(0.009)			(0.008)
Panel B. By Age Cohort						
Hispanic						
Ages 25–34	10.48	12.97	12.85	11.03	13.33	13.18
	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
Ages 50–59	10.17	12.82	12.74	10.37	12.79	12.68
	(0.04)	(0.06)	(0.04)	(0.04)	(0.06)	(0.03)
Mexican						
Ages 25–34	10.05	12.76	12.71	10.36	13.08	13.04
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Ages 50–59	8.73	12.45	12.55	8.84	12.23	12.46
	(0.05)	(0.08)	(0.05)	(0.05)	(0.08)	(0.04)
Non-Hispanic White						
Ages 25–34			13.86			14.28
			(0.006)			(0.006)
Ages 50–59			13.81			13.86
			(0.006)			(0.006)
Non-Hispanic Black						
Ages 25–34			13.01			13.33
			(0.02)			(0.01)
Ages 50–59			12.84			13.23
			(0.02)			(0.02)

Source: 2003–2019 Current Population Survey outgoing rotation group microdata from IPUMS-USA.
Notes: Standard errors are reported in parentheses. The "first generation" consists of foreign-born individuals. The "second generation" consists of US-born individuals who have at least one foreign-born parent. Remaining persons are members of the "third+ generation" (that is, the third and all higher generations), which consists of US-born individuals who have two US-born parents. Hispanics and Mexicans are identified from the Hispanic origin question in the CPS, and non-Hispanic Blacks and non-Hispanic Whites are identified from the race question. Sampling weights are employed.

resulting in substantial earnings deficits for third-plus-generation Hispanics and Mexican Americans relative to non-Hispanic Whites (Duncan and Trejo 2018).

Empirical patterns such as these have prompted lively debates among social scientists about the prospects of Mexican Americans for upward socioeconomic mobility (for example, Alba, Kasinitz, and Waters 2011; Alba, Jiménez, and Marrow 2014; Haller, Portes, and Lynch 2011a; 2011b; Park, Myers, and Jiménez 2014; Bean, Brown, and Bachmeier 2015; Ortiz and Telles 2017). Huntington (2004) voices an especially pessimistic view on this issue. He argues that several factors retard the pace of integration by Hispanic immigrants and their descendants today as compared with the European immigrants who arrived in the past. In particular, Huntington cites the size and persistence of immigration flows from Mexico and other Spanish-speaking countries, the geographic concentration of where Hispanics settle within the United States, and the ease of return and repeat migration afforded by Mexico's close proximity. According to Huntington, these unique features of Hispanic immigration discourage assimilation and instead foster the growth of Spanish-speaking enclaves where immigrants and their descendants can live and work without being forced to learn English or to Americanize in other important ways. Contrary to Huntington's thesis, however, is evidence of pervasive linguistic assimilation for Hispanics by the third generation. For example, among third-generation Mexican Americans living in Southern California, 96 percent of such individuals prefer to speak English rather than Spanish at home, and only 17 percent retain the ability to speak fluent Spanish (Rumbaut, Massey, and Bean 2006).

Moreover, the evidence of educational stagnation for Mexican Americans after the second generation, as displayed in panel A of Table 3, suffers from at least two serious shortcomings. First, cross-sectional comparisons of immigrant generations can be misleading because they do a poor job of matching cohorts of parents and grandparents in earlier generations with their descendants in later generations (Smith 2003; 2006). To address this issue, panel B of Table 3 reports separate calculations for two specific age groups: 25–34 and 50–59. Because these age groups are 25 years apart, the older age group from a particular immigrant generation potentially represents the parental cohort for the younger age group in the next generation. For example, the cohort of immigrant men ages 50–59 includes fathers of the second-generation cohort of sons ages 25–34. Under the assumption that educational attainment does not change much after age 25, comparisons between the relevant age/generation groups approximate intergenerational changes between cohorts of parents and children. In panel B of Table 3, such comparisons are made by moving northeast between the connected cells with similar shading. It is important to note that comparisons of this type between the first and second generations will more closely approximate changes between cohorts of parents and children than do the corresponding comparisons between the second and third-plus generations, because the third-plus-generation group includes some individuals who are fourth generation or higher and therefore not the child of a second-generation parent.

In panel B of Table 3, we begin to see signs of educational progress for Mexican Americans after the second generation. For men, these gains are about a quarter of a year: that is, average years of schooling of 12.71 years for the younger cohort of third-plus-generation Mexican Americans compared with 12.45 years for the older cohort of second-generation Mexican Americans. For women, the analogous gains are considerably larger, at four-fifths of a year. For both men and women, the corresponding educational improvement between first- and second-generation Mexican Americans now exceeds four years, larger than the cross-sectional improvement of slightly more than three years shown in panel A of Table 3. Still, despite these generational advances, young third-plus-generation Mexican Americans continue to trail the average schooling of their non-Hispanic White peers by more than a year. Research on educational achievement deficits for Hispanics by immigrant generation is consistent with this pattern (Schneider, Martinez, and Owen 2006). More research is needed to understand the underlying causes of this phenomenon, which may require greater availability of detailed educational data across generations.

Ethnic Attrition

A second important limitation of the evidence suggesting generational stagnation among Mexican Americans arises from the difficulty of identifying later-generation individuals. As noted previously, the first- and second-generations can be identified in the Current Population Survey (CPS) using only information about the countries of birth of the respondent and his parents (for example, second-generation Mexican Americans are US-born individuals with a parent born in Mexico). Similar to virtually all large-scale, nationally-representative US surveys, however, the CPS does not collect information about the countries of birth of the respondent's grandparents or earlier ancestors. Therefore, in order for a third-plus-generation individual to be identified as Mexican American, the individual must answer the Hispanic origin question in the following way: they must first affirm that they are "of Hispanic, Latino, or Spanish origin," and then they must indicate that the specific Hispanic group they "most closely identify with" is "Mexican" (or "Mexican American" or "Chicano").

In fact, many US-born descendants of Hispanic immigrants fail to identify as Hispanic in response to a question about Hispanic origin of the type asked in the Census and other surveys. The Pew Research Center administered surveys designed to measure such "ethnic attrition" among Hispanics (Lopez, Gonzalez-Barrera, and López 2017). The findings reveal sizeable ethnic attrition among later-generation Hispanics. The percentage of individuals with Hispanic ancestry who fail to identify as Hispanic rises sharply across immigrant generations from just 3 percent for the first generation and 8 percent for the second generation to 23 percent for the third generation and 50 percent for those who are fourth generation or higher. Analyzing 2003–2013 data from the Current Population Survey, Duncan and Trejo (2017) report similar rates of Hispanic ethnic attrition for first- and second-generation adults and third-generation children. For children living with both parents, these data can identify the third generation objectively from information about personal,

parental, and grandparental countries of birth that can be obtained by linking records between each child and his parents.

Not only is there substantial ethnic attrition among later-generation Hispanics, but this attrition is strongly selective on socioeconomic attainment (Duncan and Trejo 2007; 2011; 2017) and health (Antman, Duncan, and Trejo 2016; 2020). Both direct and indirect evidence indicate that the later-generation descendants of Hispanic immigrants who continue to self-identify as Hispanic come from less advantaged family backgrounds and have much lower levels of attainment than their counterparts who no longer self-identify as Hispanic. This pattern of substantial and selective ethnic attrition prevails for Hispanics overall, and also for Mexican Americans and other Hispanic national origin groups. As a result, analyses that rely on the Hispanic origin question or similar measures of subjective self-identification to detect third-plus-generation Hispanics will understate the socioeconomic achievement of this population. Table 3 suffers from this bias and so do almost all previous empirical studies of generational progress for Hispanics overall and for specific Hispanic national origin groups. Although available evidence makes clear the direction of this bias, data limitations have made it difficult to assess the magnitude of the bias. Nevertheless, findings from both historical (Kosack and Ward 2020; Duncan and Trejo 2022) and contemporary data (Duncan and Trejo 2011; Duncan et al. 2020) suggest that ethnic attrition produces sizeable downward bias in standard measures of attainment for Hispanics in the third generation and beyond.[6]

Intermarriage

For Hispanics and other racial-ethnic groups, intermarriage is a fundamental source of ethnic attrition. Individuals with Hispanic ancestry on both mother's and father's side almost always self-identify as Hispanic, regardless of generation (Duncan and Trejo 2007; 2011; 2017). The overwhelming majority of ethnic attrition is among individuals with mixed racial-ethnic origins. Rates of intermarriage between Hispanics and other groups rise sharply with immigrant generation, and so do rates of ethnic attrition, in part because children produced by Hispanic intermarriages have mixed racial-ethnic origins and are therefore less likely to self-identify as Hispanic. Moreover, because intermarried Hispanics and their spouses possess large advantages in education, English proficiency, and earnings relative to their counterparts in endogamous Hispanic marriages, this positive selectivity of Hispanic intermarriage drives the positive socioeconomic selectivity of Hispanic ethnic attrition, ultimately creating downward bias in standard measures of attainment for later-generation Hispanics (Duncan and Trejo 2007; 2011; 2017).

[6]Apart from self-identification in response to the Hispanic origin question, other potential indicators for individuals with Hispanic ancestry include having a Spanish surname, growing up in a Spanish-speaking household, and being able to speak Spanish. Within the Hispanic-origin population, however, these alternative indicators are strongly and inversely related to measures of assimilation such as intermarriage and English language ability. As a result, these indicators identify a relatively disadvantaged subset of Hispanics and generate downward-biased estimates of attainment, much in the same way as we have described for when Hispanic self-identification is used for this purpose.

Figure 3 shows how intermarriage rates vary with immigrant generation for Hispanics, and for comparison, Asians and third-plus-generation Blacks. These calculations employ the 2003–2019 data from the Current Population Survey previously used in Table 3. Panel A of Figure 3 displays intermarriage rates for married men ages 25–59 whose wives are at least age 18 and who live in the same household as their husband. Panel B of Figure 3 reports intermarriage rates for married women with the analogous sample restrictions. The intermarriage rates shown here represent the percent of currently married individuals whose major racial-ethnic category differs from that of their spouse. For example, a Hispanic husband with a non-Hispanic wife is defined to be intermarried, regardless of the immigrant generations of the husband and wife. The samples for Figure 3 exclude individuals with same-sex spouses.

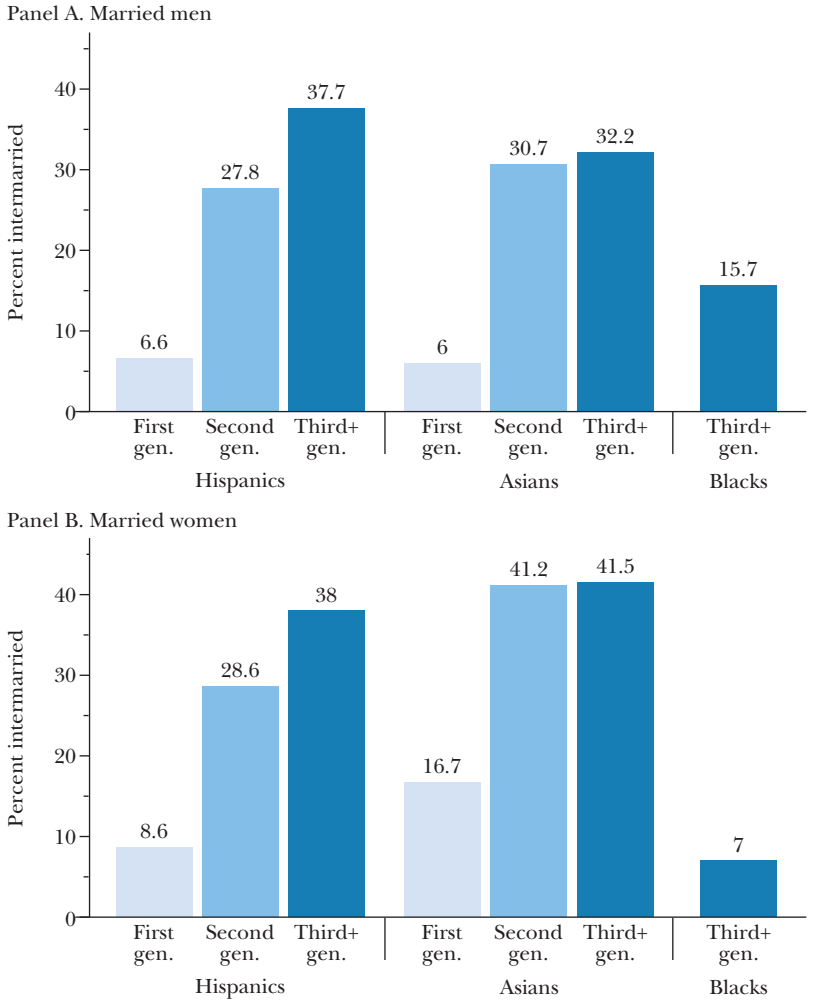
For Hispanic men, intermarriage rates increase dramatically across immigrant generations, from 7 percent for the first generation to 28 percent for the second generation and 38 percent for the third-plus generation. Within each generation, intermarriage rates for Hispanic women are similar to those for men. In part, the very low rates of intermarriage for Hispanic immigrants in Figure 3 arise because many of these individuals married before coming to the United States (Passel, Wang, and Taylor 2010).

Compared to Hispanics, intermarriage patterns differ for Asians in two ways. First, among Asians, intermarriage rates in every generation are about 10 percentage points higher for women than for men. In addition, Asian intermarriage rates rise sharply between the first and second generations but show little further increase between the second and third-plus generations. Despite these differences, intermarriage rates in the third-plus generation are of comparable magnitude for Hispanics (38 percent for both men and women) and Asians (32 percent for men and 42 percent for women). Notably, intermarriage is much more common for US-born Hispanics and Asians than it is for third-plus-generation Blacks (16 percent for men and 7 percent for women). Passel, Wang, and Taylor (2010) and Livingston and Brown (2017) provide further discussion of these and related intermarriage patterns.

Frequent intermarriage by US-born Hispanics has significant implications. Social scientists have long regarded intermarriage as a key indicator of social integration by minority racial-ethnic groups, particularly for groups with recent immigrant origins (Gordon 1964; Alba and Nee 2003), and so the relatively high rates of intermarriage by third-plus-generation Hispanics provide important evidence of such integration. In addition, by creating multiple and often weakened ethnic attachments for future generations (Hout and Goldstein 1994; Perlmann and Waters 2007), intermarriage increases ethnic attrition and the resulting biases in measuring socioeconomic progress across immigrant generations (Duncan and Trejo 2017).

Further calculations from these same data illustrate the strong positive selectivity of Hispanic intermarriage with respect to educational attainment that has been found in previous research (Duncan and Trejo 2007; 2011; 2017). For example, among married men and women who are third-plus-generation Hispanics, average

Figure 3
Intermarriage Rates, by Sex, Race-Ethnicity, and Immigrant Generation



Source: 2003–2019 Current Population Survey basic monthly microdata from IPUMS-CPS.
Notes: The intermarriage rates shown here represent the percent of currently married individuals whose major racial-ethnic category differs from that of their spouse. For example, a Hispanic husband with a non-Hispanic wife is defined to be intermarried, regardless of the immigrant generations of the husband and wife. In panel A, the samples include married men ages 25–59 whose wives are at least age 18 and live in the same household as their husband. In panel B, the samples include married women ages 25–59 whose husbands are at least age 18 and live in the same household as their wife. Both panels exclude individuals with same-sex spouses. Sampling weights are employed.

schooling is more than a year greater for those with non-Hispanic spouses than for those with Hispanic spouses. Moreover, this educational gap is even larger between the corresponding spouses—that is, between the non-Hispanic spouses of intermarried Hispanics and the Hispanic spouses of endogamously-married Hispanics.

Having better-educated parents creates educational and other advantages for the children produced by Hispanic intermarriages, and these children also are more prone to ethnic attrition than are children from endogamous Hispanic marriages. In this way, the positive educational selectivity of Hispanic intermarriage generates the corresponding positive selectivity observed for Hispanic ethnic attrition.

Recent advances in linking parents and children (and sometimes also grandparents) across historical Censuses (Kosack and Ward 2020; Ward 2020; Abramitzky et al. 2021b) and in creating similar links using more recent Census or American Community Survey data and tax records (Chetty et al. 2020; Abramitzky et al. 2021b; Lowrey et al. 2021) are very promising developments for studying progress across immigrant generations. By providing information on the countries of birth of grandparents and possibly earlier ancestors, such linkages enable estimates of intergenerational mobility for Hispanics and other immigrant groups that mitigate problems associated with selective ethnic identification. However, these data remain subject to limitations. For example, with historical Census data, cross-generation linkage rates are typically low, linked samples are not representative, and surname changes upon marriage make it difficult to link women (Bailey et al. 2020; Abramitzky et al. 2021a). Modern linkages between Census data and tax records omit undocumented immigrants and their descendants (Chetty et al. 2020; Abramitzky et al. 2021b), a particularly unfortunate omission for studying Hispanics. Despite such caveats, it is interesting to note that, using contemporary data, Chetty et al. (2020) estimate rates of intergenerational mobility (between parents and children) for Hispanics that are close to those for non-Hispanic Whites, and as a result the relative income disadvantage of Hispanics narrows across generations. In contrast, estimated rates of intergenerational mobility are much lower for Blacks and American Indians, creating income disparities for these groups that persist across generations.

Conclusion

Hispanics now comprise nearly one-fifth of the US population. One-third of Hispanics are foreign-born immigrants, and this group has attracted a large amount of attention from both researchers and policymakers, not just because of the size and rapid growth of this population, but also because most Hispanic immigrants arrive in the United States with relatively low skills and as a result earn substantially less than other Americans. On the bright side, Hispanic immigrant men have high rates of employment, and the lower earnings received by Hispanic immigrant workers seems largely to reflect their low levels of human capital (Duncan, Hotz, and Trejo 2006). Promising work in this area has uncovered other sources of earnings deficits for Hispanic immigrants. For example, recent research suggests that immigration enforcement policies have deleterious impacts on labor market outcomes for undocumented Hispanic immigrant workers (Amuedo-Dorantes and Antman 2022; East et al. forthcoming). Still, further research is needed in this area

to better understand what drives earnings differentials, and the extent to which they may reflect bias and discrimination.

This article has focused on the other two-thirds of Hispanics: those who were born in the United States and are the children, grandchildren, and later descendants of previous Hispanic immigrants. In some ways, their situation is not unlike that faced by descendants of the unskilled Italian and Irish immigrants who arrived here in large numbers a century or more ago. Confirming the findings of earlier work, recent studies using better data have documented large generational advances in education and earnings for the descendants of European immigrants that arrived in the late 1800s and early 1900s (Abramitzky et al. 2021b; Lowrey et al. 2021). Are the US-born descendants of Hispanic immigrants experiencing this same kind of upward mobility? The US-born children of Hispanic immigrants manage to erase large portions of the enormous schooling and earnings deficits that their foreign-born parents experienced relative to other Americans, but substantial disparities remain for second-generation Hispanics. Because of the size of the initial deficits faced by Hispanic immigrants, it is understandable that complete convergence does not take place in the second generation.

Do third-generation Hispanics make further progress? The answer remains murky. Few data sets allow for direct identification of the grandchildren of Hispanic immigrants, so empirical work on this issue typically must settle for using a "third-plus" generation consisting of US-born individuals who have two US-born parents and who also self-identify as Hispanic. Defined in this way, third-plus-generation Hispanics exhibit only modest gains over the second generation, and large gaps in education and earnings persist between third-plus-generation Hispanics and other Americans. However, this way of defining third-plus-generation Hispanics misses a sizeable and selective portion of the target population, because many later-generation descendants of Hispanic immigrants fail to self-identify as Hispanic. Moreover, Hispanic "ethnic attrition" disproportionately occurs among individuals with higher socioeconomic attainment. The bottom line is that later-generation Hispanics are likely doing better than previous research suggests, but how much better is hard to say at this point.

Future research would benefit from better data documenting generational status, language, race, and other differences among the large and diverse population of Hispanics. For example, skin color may affect Hispanic self-identification and be correlated with socially perceived race as well as with individuals' experiences of discrimination (Noe-Bustamante et al. 2021; Antman 2022). A very promising development, however, is that recent progress in creating data sets that link family members across generations has the potential to provide much improved estimates of the socioeconomic attainment of later-generation Hispanics.

■ *The authors thank Erik Hurst, Nina Pavcnik, Timothy Taylor, and Heidi Williams for helpful feedback. Antman acknowledges partial research support by the National Science Foundation, under NSF Award Number SES: 2121120.*

References

Aaronson, Daniel, and Bhashkar Mazumder. 2011. "The Impact of Rosenwald Schools on Black Achievement." *Journal of Political Economy* 119 (5): 821–88.

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Perez. 2021a. "Automated Linking of Historical Data." *Journal of Economic Literature* 59 (3): 865–918.

Abramitzky, Ran, Leah Boustan, Elisa Jacome, and Santiago Perez. 2021b. "Intergenerational Mobility of Immigrants in the United States over Two Centuries." *American Economic Review* 111 (2): 580–608.

Ahn, Taemin, and Rodrigo Dominguez-Villegas. 2022. "A Change of Plans: How the Pandemic Affected Students of Color and Their Plans for Higher Education." UCLA Latino Politics and Policy Initiative Policy Report, March 17. https://latino.ucla.edu/research/fact-sheet-education-covid/.

Alba, Richard D., and Victor Nee. 2003. *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Cambridge, MA: Harvard University Press.

Alba, Richard, Tomás R. Jiménez, and Helen B. Marrow. 2014. "Mexican Americans as a Paradigm for Contemporary Intra-group Heterogeneity." *Ethnic and Racial Studies* 37 (3): 446–66.

Alba, Richard D., Philip Kasinitz, and Mary C. Waters. 2011. "The Kids Are (Mostly) Alright: Second Generation Assimilation; Comments on Haller, Portes and Lynch." *Social Forces* 89 (3): 763–73.

Antecol, Heather, and Kelly Bedard. 2002. "The Relative Earnings of Young Mexican, Black, and White Women." *Industrial and Labor Relations Review* 56 (1): 122–35.

Amuedo-Dorantes, Catalina, and Francisca M. Antman. 2022. "*De Facto* Immigration Enforcement, ICE Raid Awareness, and Worker Engagement." *Economic Inquiry* 60 (1): 373–91.

Antman, Francisca. 2022. "Multidimensional Identities of the Hispanic Population in the United States." Economic Policy Institute, June 15. https://www.epi.org/anti-racist-policy-research/multidimensional-identities-of-the-hispanic-population-in-the-united-states/.

Antman, Francisca M., and Kalena Cortes. Forthcoming. "The Long-Run Impacts of Mexican-American School Desegregation." *Journal of Economic Literature*.

Antman, Francisca M., and Brian Duncan. 2021. "American Indian Casinos and Native American Self-Identification." University of Colorado Boulder Working Paper. https://spot.colorado.edu/~antmanf/Antman&Duncan_NativeAmericanIdentity&Casinos.pdf.

Antman, Francisca, Brian Duncan, and Stephen J. Trejo. 2016. "Ethnic Attrition and the Observed Health of Later-Generation Mexican Americans." *American Economic Review* 106 (5): 467–71.

Antman, Francisca M., Brian Duncan, and Stephen J. Trejo. 2020. "Ethnic Attrition, Assimilation, and the Measured Health Outcomes of Mexican Americans." *Journal of Population Economics* 33 (4): 1499–522.

Antman, Francisca M., Brian Duncan, and Stephen J. Trejo. 2023. "Replication data for: Hispanic Americans in the Labor Market: Patterns over Time and across Generations." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E183164V1.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics* 90 (2): 300–23.

Bahrampour, Tara. 2021. "Census Race Categories Increasingly Fail to Reflect How People See Themselves." *Washington Post*, September 30. https://www.washingtonpost.com/dc-md-va/2021/09/30/census-2020-race-ethnicity-categories/.

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey. 2020. "How Well Do Automated Linking Methods Perform? Lessons from US Historical Data." *Journal of Economic Literature* 58 (4): 997–1044.

Bayer, Patrick, and Kerwin Kofi Charles. 2018. "Divergent Paths: A New Perspective on Earnings Differences between Black and White Men since 1940." *Quarterly Journal of Economics* 133 (3): 1459–501.

Bean, Frank D., Susan K. Brown, and James D. Bachmeier. 2015. *Parents without Papers: The Progress and Pitfalls of Mexican-American Integration*. New York: Russell Sage Foundation.

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013.

Blau, Francine D., and Christopher Mackie, eds. 2017. *The Economic and Fiscal Consequences of Immigration*. Washington, DC: National Academies Press.

Bleemer, Zachary. 2022. "Affirmative Action, Mismatch, and Economic Mobility after California's

Proposition 209." *Quarterly Journal of Economics* 137 (1): 115–60.

Borjas, George J., 2017. "The Labor Supply of Undocumented Immigrants." *Labour Economics* 46: 1–13.

Cadena, Brian C., Brian Duncan, and Stephen J. Trejo. 2015. "The Labor Market Integration and Impacts of US Immigrants." In *Handbook of the Economics of International Migration*, Vol. 1B, edited by Barry R. Chiswick and Paul W. Miller, 1197–1259. Amsterdam: Elsevier.

Card, David, and Alan B. Krueger. 1992. "School Quality and Black-White Relative Earnings: A Direct Assessment." *Quarterly Journal of Economics* 107 (1): 151–200.

Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter. 2020. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." *Quarterly Journal of Economics* 135 (2): 711–83.

Collins, William J., and Robert A. Margo. 2006. "Historical Perspectives on Racial Differences in Schooling in the United States." In *Handbook of the Economics of Education*, Vol. 1, edited by Eric A. Hanushek and Finis Welch, 107–54. Amsterdam: North Holland.

Duncan, Brian, and Stephen J. Trejo. 2007. "Ethnic Identification, Intermarriage, and Unmeasured Progress by Mexican Americans." In *Mexican Immigration to the United States*, edited by George J. Borjas, 227–68. Chicago: University of Chicago Press.

Duncan, Brian, and Stephen J. Trejo. 2011. "Intermarriage and the Intergenerational Transmission of Ethnic Identity and Human Capital for Mexican Americans." *Journal of Labor Economics* 29 (2): 195–227.

Duncan, Brian, and Stephen J. Trejo. 2016. "The Complexity of Immigrant Generations: Implications for Assessing the Socioeconomic Integration of Hispanics and Asians." NBER Working Paper 21982.

Duncan, Brian, and Stephen J. Trejo. 2017. "The Complexity of Immigrant Generations: Implications for Assessing the Socioeconomic Integration of Hispanics and Asians." *ILR Review* 70 (5): 1146–75.

Duncan, Brian, and Stephen J. Trejo. 2018. "Socioeconomic Integration of U.S. Immigrant Groups over the Long Term: The Second Generation and Beyond." In *The Human and Economic Implications of Twenty-First Century Immigration Policy*, edited by Susan Pozo, 33–62. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Duncan, Brian, and Stephen J. Trejo. 2022. "Which Mexicans Are White? Enumerator-Assigned Race in the 1930 Census and the Socioeconomic Integration of Mexican Americans." Manuscript. Austin, TX: University of Texas.

Duncan, Brian, V. Joseph Hotz, and Stephen J. Trejo. 2006. "Hispanics in the U.S. Labor Market." In *Hispanics and the Future of America*, edited by Marta Tienda and Faith Mitchell, 228–90. Washington, DC: National Academies Press.

Duncan, Brian, Jeffrey Grogger, Ana Sofia Leon, and Stephen J. Trejo. 2020. "New Evidence of Generational Progress for Mexican Americans." *Labour Economics* 62: 101771.

East, Chloe N., Annie L. Hines, Philip Luck, Hani Mansour, and Andrea Velásquez. Forthcoming. "The Labor Market Effects of Immigration Enforcement." *Journal of Labor Economics***.**

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, and Michael Westberry. 2022. "Integrated Public Use Microdata Series, Current Population Survey: Version 10.0 [dataset]." IPUMS, Minneapolis, MN. https://doi.org/10.18128/D030.V10.0 (accessed on November 25, 2022).

Flores, Antonio, Gustavo López, and Jynnah Radford. 2017. "2015, Hispanic Population in the United States Statistical Portrait." Pew Research Center, September 18. https://www.pewresearch.org/hispanic/2017/09/18/2015-statistical-information-on-hispanics-in-united-states-trend-data/.

Gonzalez-Barrera, Ana. 2020. "The Ways Hispanics Describe Their Identity Vary across Immigrant Generations." Pew Research Center, September 24. https://www.pewresearch.org/fact-tank/2020/09/24/the-ways-hispanics-describe-their-identity-vary-across-immigrant-generations/.

Gordon, Milton M. 1964. *Assimilation in American Life: The Role of Race, Religion, and National Origins.* New York: Oxford University Press.

Gould, Elise, Daniel Perez, and Valerie Wilson. 2020. "Latinx Workers—Particularly Women—Face Devastating Job Losses in the COVID-19 Recession." Economic Policy Institute, August 20. https://www.epi.org/publication/latinx-workers-covid/.

Gratton, Brian, and Myron P. Gutmann. 2000. "Hispanics in the United States, 1850–1990: Estimates of Population Size and National Origin." *Historical Methods* 33 (3): 137–53.

Gratton, Brian, and Emily Klancher Merchant. 2015. "An Immigrant's Tale: The Mexican American Southwest 1850 to 1950." *Social Science History* 39 (4): 521–50.

Grebler, Leo, Joan W. Moore, and Ralph C. Guzman. 1970. *The Mexican American People: The Nation's Second Largest Minority.* New York: Free Press.

Guryan, Jonathan. 2004. "Desegregation and Black Dropout Rates." *American Economic Review* 94 (4): 919–43.

Haller, William, Alejandro Portes, and Scott M. Lynch. 2011a. "Dreams Fulfilled, Dreams Shattered: Determinants of Segmented Assimilation in the Second Generation." *Social Forces* 89 (3): 733–62.

Haller, William, Alejandro Portes, and Scott M. Lynch. 2011b. "Rejoinder: On the Dangers of Rosy Lenses: Reply to Alba, Kasinitz, and Waters." *Social Forces* 89 (3): 775–82.

Hout, Michael. 2012. "Social and Economic Returns to College Education in the United States." *Annual Review of Sociology* 38: 379–400.

Hout, Michael, and Joshua R. Goldstein. 1994. "How 4.5 Million Irish Immigrants Became 40 Million Irish Americans: Demographic and Subjective Aspects of the Ethnic Composition of White Americans." *American Sociological Review* 59 (1): 64–82.

Humes, Karen, and Howard Hogan. 2009. "Measurement of Race and Ethnicity in a Changing, Multicultural America." *Race and Social Problems* 1: 111–31.

Humes, Karen R., Nicholas A. Jones, and Roberto R. Ramirez. 2011. *Overview of Race and Hispanic Origin: 2010.* 2010 Census Brief C2010BR-02. Washington, DC: US Census Bureau.

Huntington, Samuel P. 2004. *Who Are We? The Challenges to America's National Identity.* New York: Simon & Schuster.

Jones, Nicholas, Rachel Marks, Roberto Ramirez, and Merarys Ríos-Vargas. 2021. "2020 Census Illuminates Racial and Ethnic Composition of the Country." US Census Bureau, August 12. https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html.

Kenney, Genevieve M., and Douglas A. Wissoker. 1994. "An Analysis of the Correlates of Discrimination Facing Young Hispanic Job-Seekers." *American Economic Review* 84 (3): 674–83.

Kosack, Edward, and Zachary Ward. 2020. "El Sueno Americano? The Generational Progress of Mexican Americans Prior to World War II." *Journal of Economic History* 80 (4): 961–95.

Krogstad, Jens Manuel. 2020. "Hispanics Have Accounted for More Than Half of Total U.S. Population Growth since 2020." Pew Research Center, July 10. https://www.pewresearch.org/fact-tank/2020/07/10/hispanics-have-accounted-for-more-than-half-of-total-u-s-population-growth-since-2010/.

Krogstad, Jens Manuel, Jeffery S. Passel, and Luis Noe-Bustamante. 2022. "Key Facts about U.S. Latinos for National Hispanic Heritage Month." Pew Research Center, September 23 https://www.pewresearch.org/fact-tank/2022/09/23/key-facts-about-u-s-latinos-for-national-hispanic-heritage-month/ (accessed on November 30, 2022).

Livingston, Gretchen, and Anna Brown. 2017. "Intermarriage in the U.S. 50 Years after Loving v. Virginia." Pew Research Center, May 18. https://www.pewresearch.org/social-trends/2017/05/18/intermarriage-in-the-u-s-50-years-after-loving-v-virginia/.

Lopez, Mark Hugo, and Ana Gonzalez-Barrera, and Gustavo López. 2017. "Hispanic Identity Fades across Generations as Immigrant Connections Fall Away." Pew Research Center, December 20. https://www.pewresearch.org/hispanic/2017/12/20/hispanic-identity-fades-across-generations-as-immigrant-connections-fall-away/.

Lowrey, Kendal, Jennifer Van Hook, James D. Bachmeier, and Thomas B. Foster. 2021. "Leapfrogging the Melting Pot? European Immigrants' Intergenerational Mobility across the 20th Century." *Sociological Science* 8: 480–512.

Lutz, Byron. 2011. "The End of Court-Ordered Desegregation." *American Economic Journal: Economic Policy* 3 (2): 130–68.

Montejano, David. 1987. *Anglos and Mexicans in the Making of Texas: 1836–1986.* Austin, TX: University of Texas Press.

Murnane, Richard J. 2013. "U.S. High School Graduation Rates: Patterns and Explanations." *Journal of Economic Literature* 51 (2): 370–422.

Neal, Derek. "Why Has Black-White Skill Convergence Stopped?" 2006. In *Handbook of the Economics of Education*, Vol. 1, edited by Eric A. Hanushek and Finis Welch, 511–76. Amsterdam: North Holland.

Noe-Bustamante, Ana Gonzalez-Barrera, Khadijah Edwards, Lauren Mora, and Mark Hugo Lopez. 2021. "Majority of Latinos Say Skin Color Impacts Opportunity in America and Shapes Daily Life." Pew Research Center, November 4. https://www.pewresearch.org/hispanic/2021/11/04/majority-of-latinos-say-skin-color-impacts-opportunity-in-america-and-shapes-daily-life/.

Office of Management and Budget. 1997. *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity.* Washington, DC: Executive Office of the President. https://www.whitehouse.gov/wp-content/uploads/2017/11/Revisions-to-the-Standards-for-the-Classification-of-Federal-Data-on-Race-and-Ethnicity-October30-1997.pdf.

Oreopoulos, Philip, and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling."

Journal of Economic Perspectives 25 (1): 159–84.

Ortiz, Vilma, and Edward Telles. 2017. "Third Generation Disadvantage among Mexican Americans." *Sociology of Race and Ethnicity* 3 (4): 441–57.

Park, Julie, Dowell Myers, and Tomás R. Jiménez. 2014. "Intergenerational Mobility of the Mexican-Origin Population in California and Texas Relative to a Changing Regional Mainstream." *International Migration Review* 48 (2): 442–81.

Passel, Jeffrey S., Wendy Wang, and Paul Taylor. 2010. "Marrying Out: One-in-Seven New U.S. Marriages Is Interracial or Interethnic." Pew Research Center, June 4. https://www.pewresearch.org/social-trends/2010/06/04/marrying-out/.

Perlmann, Joel, and Mary C. Waters. 2007. "Intermarriage and Multiple Identities." In *The New Americans: A Guide to Immigration Since 1965*, edited by Mary C. Waters and Reed Ueda, 110–23. Cambridge, MA: Harvard University Press.

Pettit, Becky. 2012. *Invisible Men: Mass Incarceration and the Myth of Black Progress*. New York: Russell Sage Foundation.

Rangel, Jorge C., and Carlos M. Alcala. 1972. "Project Report: De Jure Segregation of Chicanos in Texas Schools." *Harvard Civil Rights-Civil Liberties Law Review* 7 (2): 307–92.

Reber, Sarah J. 2010. "School Desegregation and Educational Attainment for Blacks." *Journal of Human Resources* 45 (4): 893–914.

Ruggles, Steven, Catherine A. Fitch, Ronald Goeken, J. David Hacker, Matt A. Nelson, Evan Roberts, Megan Schouweiler, and Matthew Sobek. 2021. "IPUMS Ancestry Full Count Data: Version 3.0 [dataset]." IPUMS, Minneapolis, MN.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Megan Schouweiler and Matthew Sobek. 2022. "IPUMS USA: Version 12.0 [dataset]." IPUMS, Minneapolis, MN. https://doi.org/10.18128/D010.V12.0.

Rumbaut, Rubén G. 2006. "The Making of a People." In *Hispanics and the Future of America*, edited by Marta Tienda and Faith Mitchell, 16–65. Washington, DC: National Academies Press.

Rumbaut, Rubén G., Douglas S. Massey, and Frank D. Bean. 2006. "Linguistic Life Expectancies: Immigrant Language Retention in the Southern California." *Population and Development Review* 32 (3): 447–60.

Schneider, Barbara, Sylvia Martinez, and Ann Owens. 2006. "Barriers to Educational Opportunities for Hispanics in the United States." In *Hispanics and the Future of America*, edited by Marta Tienda and Faith Mitchell, 179–227. Washington, DC: National Academies Press.

Sedmak, Todd. 2021. "Undergraduate Enrollment Declines Nearly 8% and Community Colleges Drop 15% since Fall 2019." National Student Clearinghouse Press Release, November 18. https://www.studentclearinghouse.org/blog/undergraduate-enrollment-drops-nearly-8-and-community-colleges-decline-15-since-fall-2019/.

Siegel, Jacob S., and Jeffrey S. Passel. 1979. *Coverage of the Hispanic Population of the United States in the 1970 Census: A Methodological Analysis*. Current Population Reports Special Studies P23-82. Washington, DC: US Census Bureau.

Smith, James P. 2001. "Race and Ethnicity in the Labor Market: Trends over the Short and Long Term." In *America Becoming: Racial Trends and Their Consequences*. Vol. 2, edited by Neil J. Smelser, William Julius Wilson, and Faith Mitchell, 52–97. Washington, DC: National Academies Press.

Smith, James P. 2003. "Assimilation across the Latino Generations." *American Economic Review* 93 (2): 315–19.

Smith, James P. 2006. "Immigrants and the Labor Market." *Journal of Labor Economics*, 24 (2): 203–33.

Smith, James P., and Barry Edmonston, eds. 1997. *The New Americans: Economic Demographic, and Fiscal Effects of Immigration*. Washington, DC: National Academies Press.

Smith, James P., and Finis R. Welch. 1989. "Black Economic Progress after Myrdal." *Journal of Economic Literature* 27 (2): 519–64.

Trejo, Stephen J. 1997. "Why Do Mexican Americans Earn Low Wages?" *Journal of Political Economy* 105 (6): 1235–68.

Wang, Hansi Lo. 2021. "1 In 7 People Are 'Some Other Race' On The U.S. Census. That's A Big Data Problem." National Public Radio, September 30. https://www.npr.org/2021/09/30/1037352177/2020-census-results-by-race-some-other-latino-ethnicity-hispanic.

Ward, Zachary. 2020. "The Not-So-Hot Melting Pot: The Persistence of Outcomes for the Descendants of the Age of Mass Migration." *American Economic Journal: Applied Economics* 12 (4): 73–102.

Wollenberg, Charles. 1976. *All Deliberate Speed: Segregation and Exclusion in California Schools, 1855–1975*. Berkeley and Los Angeles: University of California Press.

Journal of Economic Perspectives—Volume 37, Number 1—Winter 2023—Pages 199–222

US Immigration from Latin America in Historical Perspective

Gordon Hanson, Pia Orrenius, and Madeline Zavodny

The share of US residents who were born in Latin America and the Caribbean plateaued in the last decade or so at about 6 percent of the total US population, after a half century of rapid growth. Given how politically fraught immigration has become in the United States, this fact has received surprisingly little attention. Although smaller immigrant populations from Central and South America continue to expand, the number of US residents born in Mexico—by far the most common origin country among current US immigrants—showed little net change between 2007 and 2019. Now that the great post-1960 Latin American immigration wave has reached a mature state, we take the opportunity to reflect on its evolving characteristics, primary causes, and possible future paths.

In terms of magnitude and duration, the Latin American wave easily earns a place among the major US immigration episodes, including nineteenth- and early-twentieth-century inflows from Ireland, Germany, and Eastern and Southern Europe. As in those cases, Latin American migrants were escaping a dearth of options at home, settling initially in immigrant enclaves, and later slowly dispersing across the country (Abramitsky and Boustan 2017). Because immigration from

■ *The views expressed are those of the authors and do not reflect those of the Federal Reserve Bank of Dallas or Federal Reserve System. Gordon Hanson is Peter Wertheim Professor in Urban Policy, Harvard Kennedy School, Cambridge, Massachusetts. Pia Orrenius is Vice President and Senior Economist, Federal Reserve Bank of Dallas, Dallas, Texas. Madeline Zavodny is Professor of Economics, University of North Florida, Jacksonville, Florida. Hanson is a Research Associate, National Bureau of Economic Research, in Cambridge, Massachusetts. Orrenius and Zavodny are Research Associates, IZA Institute of Labor, Bonn, Germany. Their email addresses are gordon_hanson@hks.harvard.edu, pia.orrenius@dal.frb.org, and m.zavodny@unf.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.1.199.

the Latin American region is non-European in origin, involves populations with much less schooling than the US native-born, and includes many members who first entered the United States without authorization, there has been concern over whether large-scale inflows from the region harm US workers and deepen US cultural and political divisions (for example, Huntington and Dunn 2004). Such controversies arose with prior immigrant waves as well; the Irish were singled out for not being Protestant, the Germans for not speaking English, and Eastern and Southern Europeans for not being literate and for not being from traditional origin countries in Western and Northern Europe (Higham 2002). Calls for more immigration restrictions at that time, which culminated in the imposition of tight quotas in the 1920s,[1] have their echo in modern calls for stricter controls, which have led to more border enforcement to prevent undocumented entry (Orrenius and Zavodny 2010; Bazzi et al. 2021). Each successive influx has brought an immigrant group to the United States that at the time seemed more culturally or socially distant than the last, only for the integration of each group into American society to proceed over the ensuing decades. To date, immigration from Latin America is broadly consistent with this pattern.

Turning to migration causes, we consider how demographic shifts, economic crises, and natural disasters contributed to cross-border labor flows. We argue that, up to the COVID-19 pandemic, the long-run forces behind Latin American migrant inflows appear to have weakened, albeit unevenly, across sending countries. The acceleration of inflows in the 1980s, primarily from Mexico, was due to rapidly increasing numbers of young people entering the labor force, repeated financial crises, and a US economy that was enjoying steady growth (Hanson, Liu, and McIntosh 2017). Since then, demographic pressures for migration have slackened across Latin America (Hanson and McIntosh 2016), and, at least prior to the pandemic, economic volatility has dampened. At the same time, the US government dramatically expanded policing of US borders (Roberts, Alden, and Whitley 2013), and US economic growth slowed. Although Central America's Northern Triangle—El Salvador, Guatemala, and Honduras—has experienced considerable instability and emigration (Clemens 2021), this region accounts for just 6 percent of Latin America's population and seems unlikely to generate flows commensurate with earlier surges from Mexico, which has four times the population of the Northern Triangle and shares a land border with the United States. Much of the region appears less subject to the volatile combustibility of the 1970s, 1980s, and 1990s, when the Latin American migration wave built its momentum. Meanwhile, the challenges the United States faces regarding immigration from the region have shifted from border control to dealing with growing numbers of asylum seekers.

[1] These immigration restrictions were built on a literacy test for immigrants mandated by the Immigration Act of 1917. They included strict entry limits in the Emergency Quota Act of 1921, the permanent codification of visa allocations across origin countries based on pre-1890 immigration patterns in the Immigration Act of 1924, and legislation allowing for the deportation of immigrants without record of lawful entry in 1929 (Goldin 1994). The Western hemisphere was exempt from those quotas, reflecting the low level of public concern regarding immigration from the region early in the twentieth century.

Throughout the paper, we review some of the consequences for the United States of immigration from Latin America and the Caribbean. Looking ahead, we suspect that a long-run slowing of immigration from Latin America would create the need for adjustments in parts of the US economy, especially in labor-intensive industries in the Sunbelt and Western states. During the five-decade increase in immigration from Latin America, the United States saw a steadily expanding number of less-educated workers. From today's vantage point, it seems unlikely that coming decades will bring the same.

US Immigration from Latin America in Historical Perspective

Early US immigration flows from Latin America and the Caribbean, like their modern counterparts, were motivated by trouble at home and opportunity abroad.[2] Chileans headed to California during the Gold Rush of the 1850s; Cubans found work in Florida during the Prohibition era of the 1920s; and over one million Mexicans sought refuge in the United States during their country's revolution (1910–1920) and the ensuing decade of instability (Allende 1999; Durand, Massey, and Zenteno 2001). Seasonal migration flows from the region also have a long history. In the early twentieth century, farmers and ranchers in Texas sent contractors into Mexico to recruit agricultural workers. This practice, and the onset of World War II, brought about the US government-administered Bracero Program (1942–1964), which at its peak in the late 1950s brought 450,000 temporary farm laborers to the United States annually (Calavita 2010). Yet, permanent large-scale Latin American immigration to the United States did not begin until after 1960.

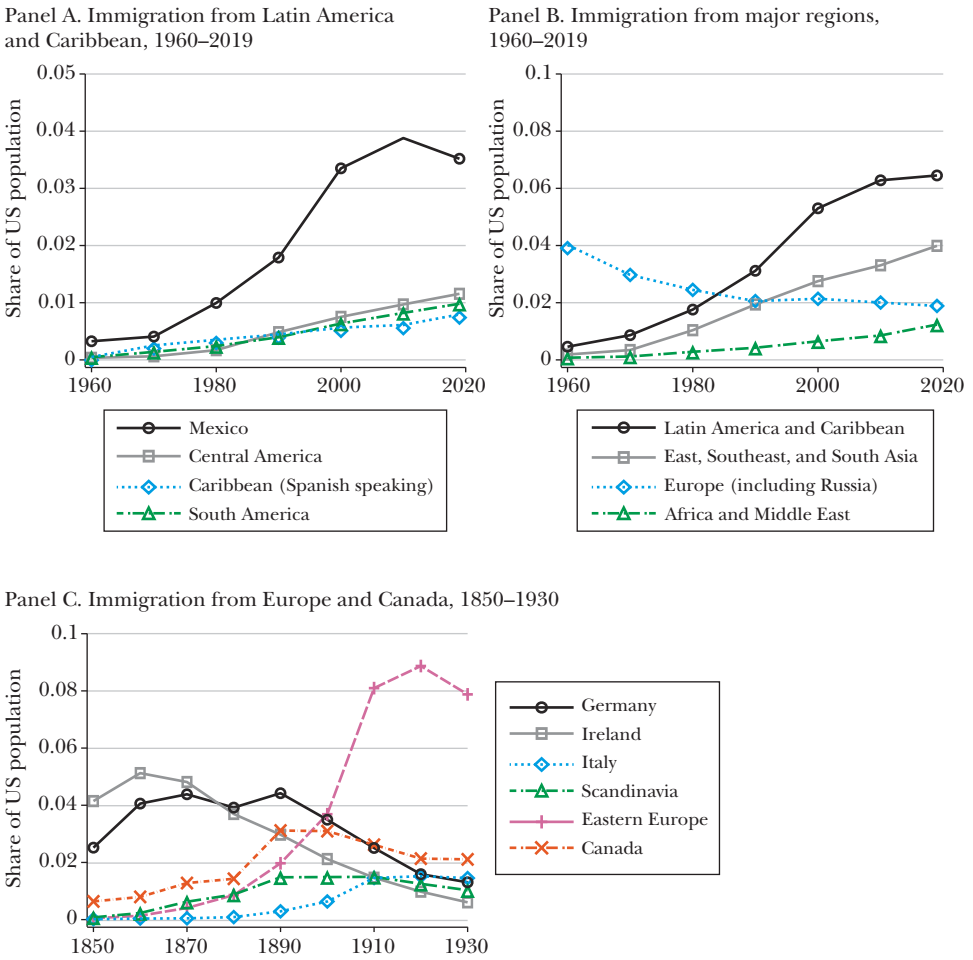
Commonalities among Major Immigration Waves

Figure 1 compares immigration from Latin America and the Caribbean to the United States from 1960 to 2019 among countries in the region (panel A), relative to other regions of origin in the same period (panel B), and relative to earlier major immigration waves (panel C). In 1960, immigrants born in Latin America were just 0.5 percent of the US population. At the time, Europe was still the largest origin region for US immigrants. Migrants from Latin America increased over time, reaching 1.8 percent of the US population in 1980, 5.3 percent in 2000, and 6.5 percent in 2019. Latin America became the top origin region of the US foreign-born population in 1990, a position it retains even though the population of immigrants from Asia grew at a faster rate during the 2010s. In 2019, immigrants from Latin America and the Caribbean were 44 percent of foreign-born residents in the United States.[3]

[2] Our discussion of immigration focuses on countries of Hispanic and Latino heritage. This includes Mexico, all of Central America (except Belize), all of South America (except French Guiana, Guyana, and Suriname), and Cuba and the Dominican Republic in the Caribbean.

[3] Since the onset of the COVID-19 pandemic in 2020, immigration from Asia has dropped sharply, whereas immigration from Latin America and the Caribbean has grown.

Figure 1
Foreign-Born Shares of the US Population



Source: Based on IPUMS data on the 1850, 1870, 1890, 1910, 1930, 1960, 1980, and 2000 US Census of Population and the 2019 1 percent sample of the American Community Survey.
Notes: In each panel and for each year, the numerator is the population of US residents from a given birth country or region and the denominator is the total US population.

Mexico is the largest source country of Latin American migrants. Its share of the US population peaked at 3.9 percent in 2010, before falling to 3.5 percent in 2019. Immigrant shares from Central America (at 1.2 percent of the US population in 2019), South America (at 1.0 percent), and the Spanish-speaking Caribbean (at 0.8 percent) are roughly similar in magnitude, despite vast differences in the respective sizes of these regions. In 2019, Cuba and the Dominican Republic together had 22 million residents, compared to 49 million in the six Central American countries,

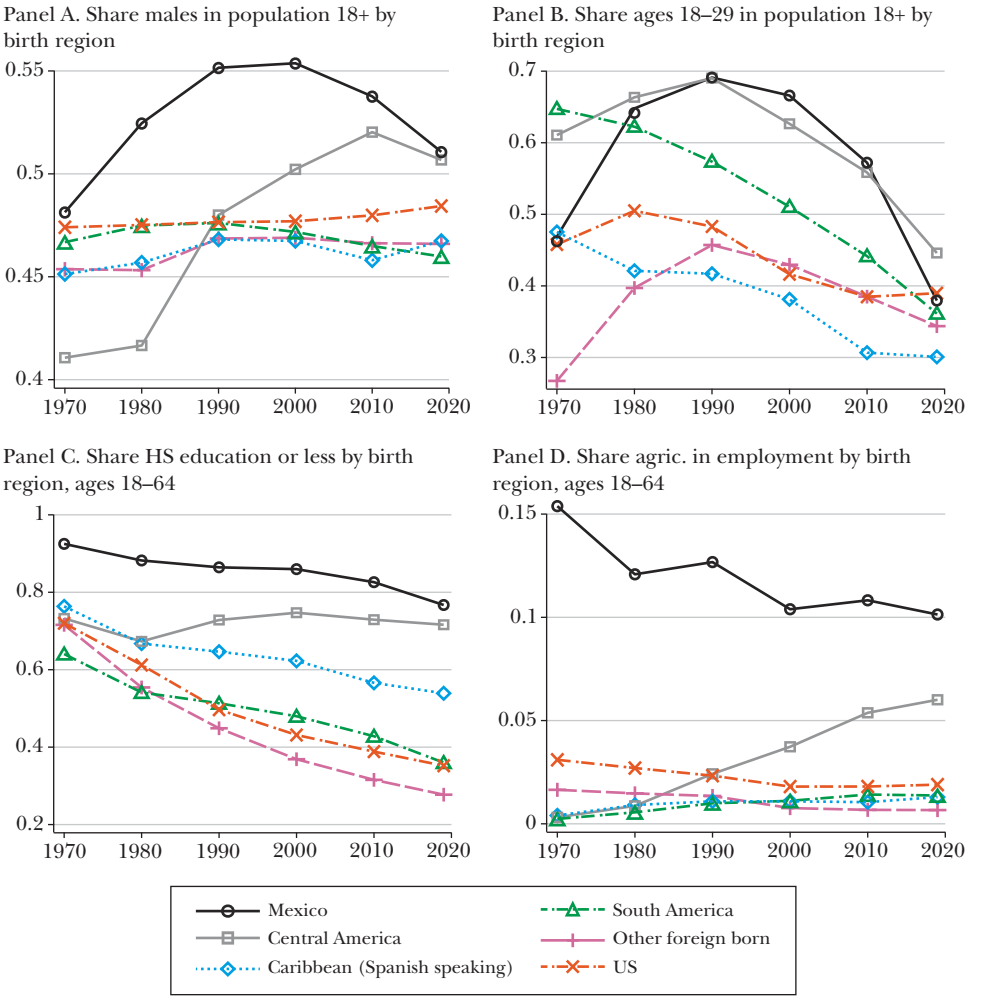
and 429 million in the nine South American countries. The implied differences in emigration rates to the United States are enormous. In 2019, origin-country immigrant populations in the United States were equivalent to 12.2 percent of Cuba's domestic population, 9.0 percent of Mexico's population, and 7.0 percent of Honduras's population, compared to just 1.7 percent of Colombia's population—which is the largest origin country for US immigrants from South America.

Looking back in US history, modern Latin American immigration is comparable in magnitude to the larger waves of the nineteenth and early twentieth centuries, as seen in Figure 1C. Because the US Census did not enumerate the birth country of respondents until 1850, the data do not fully reflect immigration from Ireland, which began in the 1820s and accelerated after the onset of the Irish Potato Famine in 1845, and from Germany, which, while most expansive after 1850, had begun earlier. Immigration from Mexico is similar in scale to inflows from these two countries, while being smaller than that from Eastern Europe in the early 1900s and larger than the respective Canadian, Scandinavian, and Italian inflows of the late 1800s. In the Irish, German, and Mexican cases, the immigrant population peaked at 4 to 5 percent of the US population and required four decades to reach this apex. Like immigrants from Mexico, who first concentrated close to the US-Mexico border, the Irish settled in Boston—the closest US port to their embarkation point of Liverpool, England—and in New York, the largest port on the US East Coast at the time (Glaeser 2005). Irish inflows were also met with political opposition, like the modern inflows from Mexico. The Know Nothing Party (1854–1856), whose platform was anti-Catholic and anti-Irish, had its greatest electoral success in Massachusetts (Alsan, Eriksson, and Niemesh 2020). In the modern era, opposition to immigration reemerged in the 1980s, contributing to the passage of the Immigration Reform and Control Act of 1986, and intensified further in the 1990s, finding notable expression in the presidential campaign of Pat Buchanan in 1992 and California's failed Proposition 187 in 1994, which sought to block undocumented immigrants from the use of all non-emergency state-level programs (Hanson 2005).

Of Sojourners and Settlers

Cuba and Mexico dominated post-1960 immigration from Latin America and the Caribbean. By 1980, the two countries accounted for nearly three-quarters of Latin American immigrants in the United States. The nature of their migrations, however, differed substantially. After Cuba's 1959 revolution, the number of Cuban immigrants in the United States increased from 78,000 in 1960 to 455,000 in 1970. Those with higher incomes, who were more at risk of being jailed or having property confiscated, were more likely to flee. As seen in Figure 2, in 1970, the US immigrant population from the Spanish-speaking Caribbean, which was overwhelmingly Cuban in origin, was modestly more female than male (panel A) and had an age distribution (panel B) and education levels (panel C) similar to the US native-born. In later years, the Cuban government occasionally permitted large-scale emigration, such as the Mariel Boatlift of 1980 (Card 1990). These later waves were representative of the broad swath of Cuban society, which has much less schooling than the US native-born, and contributed to the slower decline in the less educated among

Figure 2
Composition of Immigration from Latin America and the Caribbean



Source: Based on IPUMS data on the 1970, 1980, 1990, and 2000 US Census of Population and the 2010 and 2019 1 percent samples of the American Community Survey.
Notes: In panels (A) and (B), the population is adults (those ages 18 and up) by country or region of birth; in panels (C) and (D), the population is individuals of working age (ages 18 to 64) by country or region of birth.

immigrants from the Caribbean relative to other Latin American origin groups (panel C). In the Caribbean-origin group, Cuban migrants were later joined by migrants from the Dominican Republic, who also settled in Florida but in much larger numbers in New York City.

In the 1960s and 1970s, the characteristics of immigrants from Mexico differed from the US native-born much more than did those of immigrants from Cuba. Because of the long history of Mexican laborers travelling north to work on

US farms, much initial immigration from the country was tied to agriculture. In Figure 2, early Mexican immigrants were more likely to be male (panel A), young (panel B), and lacking a secondary education (panel C) when compared either to other Latin American immigrants or later immigrants from Mexico. In 1970, 15.4 percent of Mexican immigrants worked in agriculture (panel D), compared to 3.1 percent of US native-born workers and less than 0.5 percent of workers from elsewhere in Latin America. At that time, many Mexican workers moved back and forth across the border, following the seasonal cycle of farm jobs, while their families remained at home. This practice was viable in part because, until the 1990s, the US-Mexico border was lightly enforced (Durand, Massey, and Zenteno 2001). Migrants without visas could cross the border with little consequence and with success likely within several attempts. Over time, these sojourners became settlers (Marcelli and Cornelius 2001). Mexican immigrants spread beyond agriculture and included more women. The expansion of US border enforcement—first in the early 1990s after the Immigration Reform and Control Act of 1986, and then in the 2000s after the terrorist attacks of September 11, 2001—made circular migration riskier and costlier (Gathmann 2008). In response, more Mexican immigrants chose to reside in the United States on a permanent basis (Angelucci 2012).

Immigration from Central and South America expanded after 1980 during periods of economic and political volatility in the region. The Northern Triangle countries of El Salvador, Guatemala, and Honduras dominated flows from Central America and accounted for 85 percent of the US population from that subregion in 2019. Other countries in Central America include relatively prosperous Costa Rica and Panama, which send few migrants to the United States, and relatively poor Nicaragua, most of whose sizable emigrant population resides in neighboring Costa Rica. Whereas from the 1970s to the 2000s migrants from Mexico accounted for the vast majority of those apprehended trying to cross the US-Mexico border without authorization, by the mid-2010s apprehensions of migrants from Northern Triangle countries had become roughly equal to those from Mexico (US Department of Homeland Security 2022).

Turning to South America, the largest origin countries for US immigrants are (in descending order of their 2019 immigrant populations) Colombia, Brazil, Peru, and Ecuador, which together accounted for 71 percent of immigrants from the subregion in 2019. Much emigration from South America has not been northward to the United States so much as within the continent or to former colonial powers, such as Spain, which at times has allowed the entry of Latin Americans without a visa. For example, most Ecuadorians who left during an economic crisis in the late 1990s went to Spain, while Venezuelans who exited as their economy collapsed in the mid-2010s primarily went to Colombia (Bertoli 2010; Wolfe 2021).

The Pandemic Interregnum

Following the onset of the COVID-19 pandemic in 2020, the immigration trends of the preceding decade partially reversed. Notably, attempted unauthorized immigration from Latin America and the Caribbean soared. US Border Patrol encounters with unauthorized migrants at the US-Mexico border rose from 1 million in Fiscal

Year 2019 to 2.4 million in Fiscal Year 2022, with the large majority of these migrants coming from Latin America and the Caribbean.[4] The migration surge came on the heels of widespread COVID-19 restrictions and severe economic downturns in Latin America. Rising attempts at undocumented immigration after origin-country crises are a familiar pattern (Hanson and Spilimbergo 1999).

In about half of migrant encounters—primarily involving adults from Mexico and the Northern Triangle countries—the US Border Patrol summarily expelled those apprehended under Title 42 of the US Code, which allows the government to prohibit migrant entry during a public health emergency in order to avoid the spread of disease. From April 2020 to March 2022, migrants from Mexico accounted for 60 percent of Title 42 expulsions, while migrants from the Northern Triangle accounted for another 34 percent of expulsions (as reported by Gramlich 2022). Many expelled migrants reattempted entry and were caught again, inflating the number of encounters (Bazzi et al. 2021). Although counts of migrant encounters along the border are available, we do not yet know how the US stock or flow of undocumented immigrants changed during the pandemic.

Most of the remaining pandemic-era migrant encounters at the US-Mexico border have involved people seeking asylum. The US immigration system has historically allowed people to present themselves to authorities at a US border, request US admission as an asylum seeker, and remain in the United States until their asylum claim is adjudicated.[5] Unauthorized immigrants who are apprehended in the United States can also seek asylum as a defense against deportation. Given the rise in asylum-seekers even before the pandemic, the United States had stopped allowing most of these individuals to enter and instead required them to wait in Mexico. The backlog of asylum claims has grown rapidly, and it typically takes years for an applicant to go through the asylum claim process. It is unclear how the US government will resolve the backlog or whether it will continue to allow applicants to remain in the United States while they await adjudication of their cases.

Causes of Immigration from Latin America and the Caribbean

Immigration from Latin America and the Caribbean started gradually in the 1960s, grew at an increasing rate from 1970 to 2000, and then rose at a decreasing rate from 2006 to 2019, as shown earlier in Figure 1. This pattern reflects the timing of the shocks that contributed to labor outflows from the region, the internal forces that sustained migration once it had initiated, and the increasing restrictiveness of US immigration policy.

[4] The figures are reported at the US Customs and Border Protection website at https://www.cbp.gov/newsroom/stats/southwest-land-border-encounters.

[5] The rules are discussed at the US Citizenship and Immigration Services website at https://www.uscis.gov/humanitarian/refugees-and-asylum/asylum.

The Decision to Migrate

In modelling migration, economists posit that individuals weigh the benefits and costs of moving. Benefits of migration include the possibility of earning higher wages abroad, escaping violence or political repression at home, and achieving a better future for one's children. Clemens, Montenegro, and Pritchett (2019) compare the average earnings of young foreign-born men with a secondary education who moved to the United States to those who stayed in their birth country. Among those born in Latin America, the ratio of US to origin-country earnings in 2000 (adjusted for purchasing power parity) ranged from lows of 2.1 for the Dominican Republic and 2.6 for Mexico to highs of 3.8 for Brazil and 4.2 for Peru. (Values for Cuba and the Northern Triangle countries of Central America were not available.) Purely in terms of real earnings, the gains from migration appear substantial.

On occasion, the benefits from migrating rise suddenly due to a deterioration in origin-country conditions brought on by economic crises, natural disasters, or political upheaval. Beyond the Cuban Revolution of 1959, currency collapses in Mexico in 1982 and 1994 and several devastating hurricanes elsewhere in Latin America triggered substantial outflows (Mahajan and Yang 2020; Monras 2020). In the Mexican case, apprehensions of those crossing the US-Mexico border illegally—a proxy for undocumented immigration—showed large and rapid responses to exchange-rate-induced changes in US-Mexico relative wages during the 1980s and 1990s (Hanson and Spilimbergo 1999).[6] Geographic proximity to the United States meant that adverse shocks translated quickly into cross-border labor flows. Meanwhile, US GDP grew steadily (at least relative to Mexico's GDP) during the Great Moderation of 1982 to 2007, creating a continuing lure to prospective migrants experiencing volatility in Latin America. Net migration from Mexico came to an abrupt halt with the onset of the Great Recession in the United States in 2007. Economic contractions in much of Latin America during the COVID-19 pandemic combined with sharply higher wages in the US likely increased pressures to emigrate. In Cuba, Nicaragua, and Venezuela, greater political repression may have compounded these pressures.

Other important causes of migration are slower moving. Over time, demographic shifts may alter relative labor supplies, and therefore relative wages, across countries. In the 1970s and 1980s, Latin America and the Caribbean began to see relatively large cohorts of young adults entering the labor market, which in theory should have put downward pressure on domestic wages (Hanson and McIntosh 2012). In Mexico, the total fertility rate reached a stunning seven births per woman in the mid-1960s, which meant record growth in labor supply two decades later (Hanson and McIntosh 2009). When repeated economic crises hit Mexico in the 1980s and 1990s, these demographic-induced downward pressures on wages helped push migrants abroad. Drug-related violence is an additional slow-moving cause of migrant outflows (Orozco-Aleman and Gonzalez-Lozano 2018; Clemens 2021).

[6]The US wage expressed in terms of the Mexican peso also affected border apprehensions, suggesting that migrants planned to keep links with origin communities, whether through remittances to family members or return migration.

Barriers to Migration

The costs to migration include the financial expense of moving to the United States and the psychic burden of leaving home. Migrant networks in the destination country can help lower perceived migration costs and boost future outflows. As the stock of prior migrants from an origin country grows, new migrants may have an easier time of landing a job, finding housing, and locating places to socialize. Empirically, networks elevate the probability of migration by improving labor market outcomes for new arrivals (Munshi 2003; Orrenius and Zavodny 2005). These networks—which may be based on kinship, friendship, or simply sharing a common origin community (Caballero, Cadena, and Kovak 2018)—can make migration self-reinforcing. Because current migration lowers future migration costs, migration may continue to rise even after initial push factors have waned (Carrington, Detragiache, and Vishwanath 1996).

The costs of migrating to the United States depend on the mode of entry. Most immigrants from Latin America and the Caribbean appear to have entered the United States either without authorization or with visas sponsored by family members already in the country (Jasso et al. 2008). Unauthorized inflows grew following the end of the Bracero Program and passage of the Immigration and Naturalization Act of 1965, which imposed a cap on legal immigration from the Western hemisphere for the first time and allocated most permanent resident visas (green cards) to family members of US citizens and legal permanent residents (Massey and Pren 2012).

Most Latin American immigrants residing in the United States without authorization entered the country by crossing the US-Mexico border illegally or by obtaining a temporary visa and staying beyond its expiration (Warren 2019). Of the estimated 8.1 million undocumented immigrants from Latin America and Caribbean in the United States in 2017, 84 percent were from Mexico and Central America, while 16 percent were from South America and the Caribbean (Passel and Cohn 2019). In 2019, the respective shares of these two subregions in the overall Latin American immigrant population were 73 percent and 27 percent, indicating that Mexico and Central America are overrepresented among the region's undocumented immigrants.

By the 1990s, networks of Mexican immigrants in the United States were firmly in place. In the previous decade, the Immigration Reform and Control Act of 1986 had started a process that ultimately granted legal permanent residence to over two million undocumented immigrants from Mexico, allowing those migrants to sponsor relatives abroad for green cards—yet undocumented immigration continued (Orrenius and Zavodny 2003). Because of backlogs for visas, which are subject to annual quotas for all family members who are not immediate relatives of US citizens, many Mexican immigrants who had applied for a green card still entered the United States without authorization while they awaited adjudication of their application (Massey, Durand, and Malone 2003).

The intensification of US border enforcement starting in the 1990s has made illegal entry much more difficult. From the mid-1990s to the late 2000s, the United States quintupled the number of Border Patrol agents stationed at the US-Mexico

border, built 700 miles of physical barriers along the border, expanded legal sanctions for those caught crossing illegally, and increased the deportation of undocumented immigrants residing in the US interior (Roberts, Alden, and Whitley 2013). These changes, plus the Great Recession of 2007–2009 and the sluggish US recovery that followed, combined to reduce inflows of undocumented immigrants (Gathmann 2008; Allen, de Castro Dobbin, and Morten 2018; Lessem 2018; Bazzi et al. 2021). Between 2007 and 2019, Mexico's net migration rate to the United States turned negative, reflecting both reduced in-migration and increased voluntary and involuntary return migration (Gonzalez-Barrera 2017).

The intensification of immigration enforcement has made the pandemic-era increase in apprehensions at the US-Mexico border difficult to interpret. On the one hand, rising border apprehensions imply more people are attempting to enter illegally; on the other hand, more apprehensions may mean that, relative to the past, repeat apprehensions of migrants have increased.[7] A further source of uncertainty about recent immigration inflows is the unresolved disposition of the many Latin Americans who have applied for asylum and who remain in the United States while awaiting an immigration hearing. It will thus be some time before we know whether and by how much immigration from Latin America increased during the special period of immigration procedures instituted under the pandemic.

Selection into Immigration

From 1970 to 2019, the difference in the share of the working-age population with a high-school education or less between Mexican immigrants and the US native-born doubled from 21 percentage points (93 versus 72 percent) to 42 percentage points (77 versus 35 percent). This overall pattern of large and rising gaps in average schooling between the US native-born and immigrants from Latin America and the Caribbean is apparent in Figure 2.

One reason for this pattern is that post-secondary educational attainment is much higher in the United States than in most of Latin America.[8] It also bears noting that Mexican immigrants in the United States are drawn disproportionately from the middle of Mexico's educational distribution—they are not strongly negatively or positively selected in terms of schooling (Chiquiar and Hanson 2005).

Although Mexico has higher educational attainment than Central America or much of the Caribbean and South America, Mexico sends immigrants to the United

[7] When the US Border Patrol began expelling unauthorized migrants under Title 42 of the US Code in 2020, it stopped pursuing legal penalties against those migrants, removing an important deterrent. The share of apprehensions involving repeat crossers rose from 7 percent in the fiscal year before the pandemic to 24 percent during the pandemic (Gramlich 2022). This is suggests when the United States began imposing such penalties in 2007, it resulted in decreased recidivism in apprehensions (Bazzi et al. 2021).

[8] Between 1970 and 2010, the fraction of the population ages 15 to 64 with some post-secondary education increased from 2.2 percent to 17.8 percent in Mexico and from 22.2 percent to 55.6 percent in the United States (based on the Barro-Lee Educational Attainment Dataset available at http://www.barrolee.com/).

States who are less educated than arrivals from the other subregions.[9] This pattern arises because immigrants from everywhere else in Latin America are positively selected in terms of schooling—that is, those with more education are more likely to migrate abroad (Grogger and Hanson 2011). For these countries, migration costs to the United States are also relatively high. For example, Central Americans migrating to the United States without authorization must traverse Mexico, which involves physical risks and large smuggling fees; those from the Caribbean must cross by sea or obtain an entry visa of some kind; and those from more distant South America face greater logistical challenges still. Empirically, the higher the migration costs, the lower is the fraction of less-educated and lower-income individuals among those who emigrate (Orrenius and Zavodny 2005; McKenzie and Rapoport 2007). We thus tend to see greater positive selection of Latin American immigrants in terms of education the farther a country is from the US border.

Determinants of Migration Rates

To study the factors behind recent immigration from Latin America and the Caribbean to the United States more formally, we use data on the decadal change in the number of foreign-born from each country in the region living in the United States relative to the origin country's population at the start of each decade. We include the 18 countries from the region that had reasonably large samples in the decennial Census (1960 to 2000) and the American Community Survey (2010 and 2019); we include all ages, because a growing share of migrants from the region are children or are middle aged and beyond.

Motivated by our discussion above, we focus our regression analysis on variables that capture migrant networks and key demographic, economic, and other push factors in the region. We measure migrant networks with an indicator variable equal to one if, at the start of the decade, the number of migrants living in the United States as a share of the origin country's population is in the top half of the Latin American sample. We capture demographic pressures using the share of the origin country's population that is between ages 5 and 14 at the start of the decade, which indicates the relative size of the population that will come of working age by the end of the decade. We characterize economic push factors using the growth rate of real GDP per capita and the number of balance-of-payments crises during the decade. We distinguish between decades when GDP grew and those when it contracted, as the effects on migration may be asymmetric. We measure balance-of-payments crises using "sudden stops" in inflows of international capital or large declines in a country's current account, as documented by Cavallo (2006). We include the total number of major natural disasters, based on data from the International Emergency Event Database on the number of hurricanes, earthquakes, volcanic eruptions, floods, and droughts during the decade, and defining a natural disaster as "major" if it affected

[9] In 2010, and among the population ages 15 to 64, the 17.8 percent of Mexicans with some post-secondary education compared to 9.8 percent in Brazil, 12.2 percent in the Dominican Republic and 5.3 percent in El Salvador (again, based on the Barro-Lee Educational Attainment Dataset at http://www.barrolee.com/).

at least 10 percent of the population or killed at least 0.01 percent of the population. We also include the number of major armed conflicts during the decade using data from the Uppsala Conflict Data Program, where an armed conflict is defined as "major" if it resulted in more than 1,000 deaths since it began (and where we include both conflicts that involve the state and that involve non-state actors only).

Each of the first four columns in Table 1 presents the results of a separate regression. The dependent variable in each regression is the change in the number of foreign-born living in the United States over a decade as a share of the population in an origin country at that start of that decade. Each of the regressions also includes origin country and decade fixed effects to control, first, for time-invariant migration push factors that are specific to an origin country and, second, for pull factors that are common across all countries during a given decade. The former encompasses factors such as distance from, linguistic similarity to, and shared colonial history with the United States, while the latter absorbs the stage of the US business cycle and the intensity of US immigration restrictions.

The specifications in Table 1, columns 2 through 4, each include an interaction of the variable noted at the top of the column with a variable measuring distance (population-weighted) between the origin country and the United States. The intuition here is that distance is a proxy for bilateral migration costs, and the interaction term seeks to capture the relationship between an individual regressor and these costs.

Overall, the evidence in Table 1 is consistent with economic crises leading to migrant outflows. Countries have larger outflows to the United States during decades of economic weakness, especially as captured by the number of balance-of-payments crises. Having a balance-of-payments crisis during the decade is associated with a 0.8 to 1 percentage point increase in the decadal migration rate, roughly equivalent to the weighted sample mean of 0.8 and thus suggesting that a balance-of-payments crisis doubles outflows. The rate of GDP growth does not have a significant effect on migration from the region, whereas a higher rate of GDP contraction spurs additional migration. Although the results suggest that crises, not economic growth, lead to more migration from the region, it is important to consider that many residents are very poor and simply do not have the resources to migrate. Economic growth that leads to higher income and savings could enable more people from Latin America to undertake the costly move to the United States (Clemens 2022).

Migrant networks and origin demographics matter, and seem to matter considerably more when distance from the United States is taken into account. Being farther away dampens the positive impact of migrant networks or a youth bulge on migration, as indicated by the negative estimated coefficient on the interaction term in columns 2 and 3. Meanwhile, the results suggest that a country with its population centered 2500 kilometers from the United States—the distance between the population-weighted centers of Mexico and the United States—would see an additional 1.7 percent of its population migrate to the United States over a decade if its migrant network is in the top half of the sample (column 2).[10] A one-standard

[10] The estimate is based on evaluating the estimated coefficient on the interaction term in column 2 at 2.5 and adding the estimated coefficient on the migrant network variable.

Table 1
Determinants of US immigration rates, Latin American countries, 1960s to 2010s

		Log distance from US ×			
	(1)	*Network* (2)	*Pop. age 5–14* (3)	*Conflict* (4)	*Sample mean* (5)
Number of balance-of-payments crises during decade	1.013	1.038	0.902	0.823	0.403
	(0.496)	(0.496)	(0.451)	(0.390)	(0.638)
Rate of real GDP per capita growth over decade	–0.002	–0.002	–0.001	–0.001	31.433
	(0.006)	(0.006)	(0.006)	(0.006)	(29.053)
Rate of real GDP per capita contraction over decade (absolute value)	0.017	0.017	0.018	0.017	2.619
	(0.004)	(0.004)	(0.003)	(0.004)	(12.767)
Migrant network in top half of LACs at start of decade	0.652	3.033	0.846	0.815	0.389
	(0.452)	(1.507)	(0.441)	(0.405)	(0.490)
Share of population ages 5–14 at start of decade	0.007	–0.011	0.207	0.014	22.647
	(0.081)	(0.077)	(0.082)	(0.056)	(3.850)
Number of major natural disasters during decade	0.406	0.395	0.382	0.220	0.369
	(0.234)	(0.228)	(0.221)	(0.208)	(0.555)
Number of years with major armed conflicts during decade	–0.054	–0.070	–0.034	–0.485	1.533
	(0.076)	(0.074)	(0.061)	(0.217)	(2.826)
Interacted variable	—	–0.515	–0.034	0.092	5.807
		(0.277)	(0.008)	(0.036)	(2.492)
R^2	0.734	0.741	0.761	0.770	
Number of observations	107	107	107	107	107

Source: See the online Appendix for data sources.
Notes: Columns 1–4 report separate regressions; column 5 reports the weighted sample mean (and standard deviation) of the indicated regressor. The dependent variable is the change in the number of foreign-born living in the United States over the decade as a share of the population in the origin at that start of decade (weighted sample mean is 0.838). The sample covers 18 Latin American countries from 1960 to 2010. All specifications include country and decade fixed effects. Observations are weighted by the origin population at the start of the decade. Standard errors in parentheses in columns 1–4 are clustered on the origin country.

deviation increase in the origin population share ages 5–14 is associated with a 0.04 percentage point increase in the decadal migration rate, about one-twentieth of the mean rate, when evaluated at the average distance for the sample (column 3).[11]

Major natural disasters are an additional push factor. Having such an event is associated with a roughly 50 percent increase in the decadal migration rate (column 1). Civil conflict also appears to affect migration. Evaluated at mean distance to the United States, experiencing one year of major armed conflict is associated with a 0.05 percentage point increase in the decadal migration rate, less than one-tenth of the sample mean (column 4).[12] Being farther away from the United States implies a larger impact of armed conflict on migration flows (that is, the interaction term is positive). This surprising result may reflect heterogenous effects of violence on migration flows across origin countries. The literature reaches mixed

[11] The estimate is based on adding the estimated coefficients on the interaction term in column 3 (evaluated at mean distance and the standard deviation for the youth population share) and on the youth population share (evaluated at the standard deviation for the youth population share).
[12] The estimate is based on evaluating the estimated coefficient on the interaction term in column 4 at 5.807 (mean distance) and adding the coefficient on the armed conflicts variable, which results in a positive estimate despite the negative estimated coefficient on the main effect for the conflict variable.

findings about whether higher levels of violence cause migration (for example, Orozco-Aleman and Gonzalez-Lozano 2018). Most Latin American migration in response to natural disasters and conflict is internal, which is less costly.

Patterns of Integration

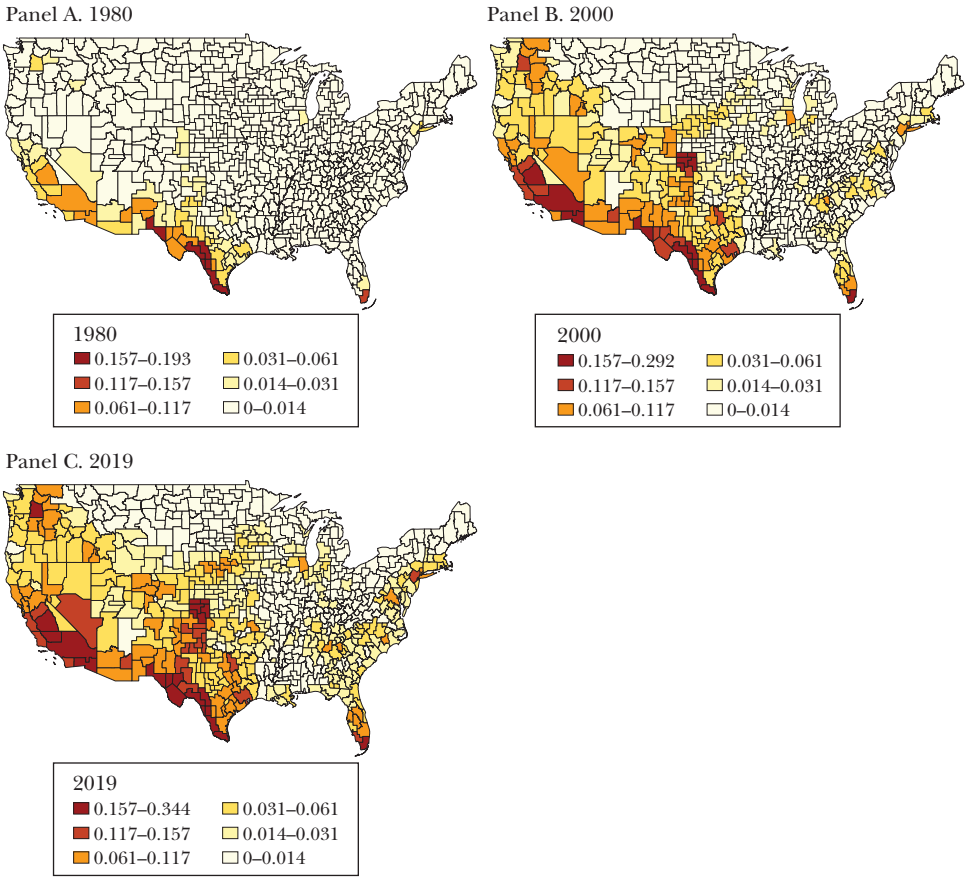
Latin American immigrants, like earlier immigrant groups, face many challenges in adapting to life in a new country, including learning a new language and customs. Large numbers of Latin American migrants have an additional challenge: lack of legal status. Many of the immigrants from the region who entered the United States without authorization have not succeeded in obtaining a green card, which creates uncertainty about their future opportunities to remain in the country. During the Age of Mass Migration from Europe in the late eighteenth and early nineteenth centuries, which occurred without the legal complications of today, the assimilation of many immigrant groups was considered slow, often stretching well into the second generation (Abramitsky et al. 2014). In this section, we examine markers of integration among Latin American immigrants related to settlement patterns, language, and citizenship.

Geographic Dispersion

In the presence of migrant networks, new immigrant arrivals in a country are likely to settle in enclaves comprised of individuals from their birth region. The concentrations of Cubans in Miami, Mexicans in Los Angeles, and Dominicans in New York City are a few of many such examples. Figure 3 describes the geographic dispersion of US immigrants from Latin America and the Caribbean. We map the share of immigrants from the region in the total population of each commuting zone for the continental United States. In 1980, when large-scale immigration from the region was barely a decade old, migrant populations were concentrated in communities close to the US-Mexico border, where Mexican immigrants tended to settle; South Florida, where Cuban immigrants tended to settle; nascent enclaves around New York City, consisting mostly of immigrants from the Caribbean and South America; and select agricultural regions in the West, here too consisting mostly of immigrants from Mexico. By 2000, in contrast, immigrant populations had spread, creating new clusters in growing urban areas, including Atlanta, Boston, Charlotte, Chicago, Dallas–Fort Worth, Denver, Houston, Raleigh-Durham, and Washington, DC. New clusters were also present in Missouri and Nebraska, where immigrants from Latin America helped fill openings in beef and pork packing plants (Champlin and Hake 2006). Between 2000 to 2019, Latin American immigrant populations grew intensively in and around the clusters that had formed by 2000 and spread only modestly beyond them.

Two factors likely contributed to the geographic dispersion of Latin American immigrants after 1980. The first is the legalization of undocumented migrants that was part of the 1986 Immigration Reform and Control Act (Orrenius and Zavodny 2003; Card and Lewis 2007). Legalization may have lowered the perceived costs of

Figure 3
Share of Commuting Zone Population Born in Latin America and Caribbean



Source: Based on IPUMS data on the 1980 and 2000 US Census of Population and the 2019 1 percent sample of the American Community Survey.
Notes: Figures show the share of the US population in a given commuting zone and year that was born in Latin America and the Caribbean. The legends divide population shares into six categories by value for the bottom four quantiles and the top two deciles.

internal migration for those who had previously lacked a green card. A second factor relates to the potential for immigrant workers to "grease the wheels" of the labor market (Borjas 2001). Because immigrants may have weaker long-run attachments to specific US cities than do the native-born, they may be more mobile in response to labor market shocks. During the Great Recession, recent Mexican immigrants with a high-school education or less were highly responsive to changes in local labor demand, whereas less-educated native-born workers were not (Cadena and Kovak 2016). This responsiveness, more generally, may have made Latin American immigrants relatively likely to move into growing US cities in the 1980s, 1990s, and early 2000s.

Language, Citizenship, and Permanence

In Figure 4, we consider additional markers of immigrant integration. Perhaps the simplest is language. In panel A, we show the fraction of the adult population that speaks English "well," "very well," or "only" by birth region. Although immigrants from Latin America are less likely to speak English than immigrants from other regions, English-speaking rates are high and stable over time at around 92 percent for South Americans and around 82 percent for those from the Caribbean, while for Mexican immigrants they have risen over time from 76 percent in 1980 to 85 percent in 2019. For Central Americans, English-speaking rates have fallen, which may reflect the recency of these flows as well as falling educational attainment among recent immigrants relative to earlier arrivals from Central America.

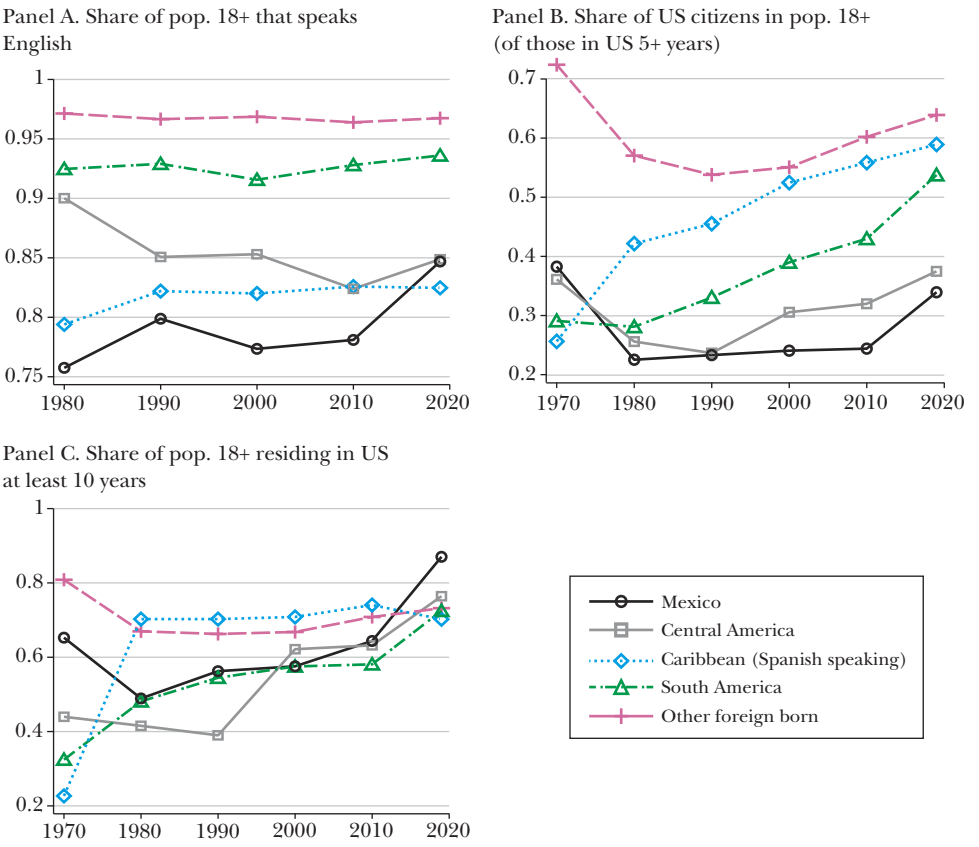
In panel B, we turn to the fraction of the immigrant population that has US citizenship, which is one indicator of being permanently attached to a country. Because immigrants with a legal permanent residence visa typically have to wait five years before they can apply for citizenship, we show citizenship rates for the population that meets this residence criterion. Not surprisingly, given the relatively high fraction of Latin American immigrants who never obtain a green card, citizenship rates for this population are lower than for immigrants from outside the region. Citizenship rates are highest for those from the Caribbean (58 percent in 2019), next highest for South Americans (54 percent), and lowest for Central Americans (38 percent) and Mexicans (34 percent). These rates reflect variation in the incidence of undocumented status among these groups. However, even among Mexicans eligible for citizenship, naturalization rates are lower than for other groups (Gonzalez-Barrera 2017). In interviews, migrants cite inadequate English skills (which make it hard to pass the citizenship test) and the cost of applying for citizenship as deterrents to naturalizing.

In panel C, we consider a third indicator of the attachment of Latin American immigrants to the United States: the fraction of adult immigrants who have resided in the country for at least ten years. In the absence of return migration (and in the presence of stable emigration rates), this fraction would rise mechanically over time. As immigration continues, new arrivals would tend to account for a smaller share of the origin group population. Among all Latin American immigrant groups, the fraction of the population with at least ten years of residence in the United States has increased over time. In 2019, it ranged from 70 percent for immigrants from the Caribbean to 87 percent for immigrants from Mexico. Based on Figure 4, there is little reason to believe that most noncitizens from Latin America might ultimately choose to return permanently to their birth country, despite their legal status being unresolved.

Immigrant Employment Patterns

Given the concentration of immigrants from Latin America and the Caribbean in specific US regions, seen in Figure 3, and their overrepresentation among those with a high school education or less, seen in Figure 2, we would expect immigrants from the region to account for a large share of employment in labor-intensive sectors. In Figure 5, we show, by US commuting zone, the 2019 employment share

Figure 4
Assimilation of the Population 18+ by Birth Region



Source: Based on IPUMS data on the 1980, 1990, and 2000 US Census of Population and the 2010 and 2019 1 percent samples of the American Community Survey.
Notes: Figures show for immigrants 18 and older from each origin country or region the share of the population that speaks English, the share of the population that is a US citizen, and the share of the population that has been residing in the United States for at least 10 years.

of Latin American immigrants in four large sectors in which less-educated workers predominate: agriculture, construction, manufacturing, and personal services.

Nationally, Latin American immigrants are a major presence in these sectors. In 2019, they accounted for 28 percent of employment in agriculture (up from 2 percent in 1970), 21 percent of employment in construction (up from 1 percent in 1970), 15 percent of employment in personal services (up from 2 percent in 1970), and 9 percent of employment in manufacturing (up from 2 percent in 1970).[13] In

[13] These shares are higher when considering less-educated workers. In 2019, the shares of Latin American immigrants in the employment of workers with a high school education or less were 42 percent in agriculture, 30 percent in construction, 24 percent in personal services, and 16 percent in manufacturing.

the regions where Latin American immigrants have concentrated, their presence is especially pronounced. At the 90th percentile of commuting zones in terms of the employment of workers born in Latin America, their employment shares are nearly 60 percent in agriculture, over 40 percent in construction, nearly 30 percent in personal services, and over 20 percent in manufacturing.

What will happen to the US labor market in the future if immigration from Latin America continues to moderate? For tradable goods production, such as in agriculture and manufacturing, firms may need to reduce labor intensity by altering product mixes or production techniques. Alternatively, firms may shift production offshore where possible. In the past, manufacturing plants that were located near US metropolitan areas experiencing larger inflows of less-educated immigrants were slower to increase machinery per unit of output (Lewis 2011), which is consistent with pressures for automation being responsive to immigration. For nontradable sectors, such as construction and personal services, relative prices may rise, and some US workers may be able to command better wages and working conditions. In the past, US local labor markets with larger inflows of less-educated immigrants had smaller increases in the relative prices of non-traded services—such as childcare, gardening, and housekeeping—than did other regions (Cortes 2008). These regions in turn saw greater displacement of native-born workers employed in occupations tied to these industries, but no such adjustment for jobs in tradable industries (Burstein et al. 2020).

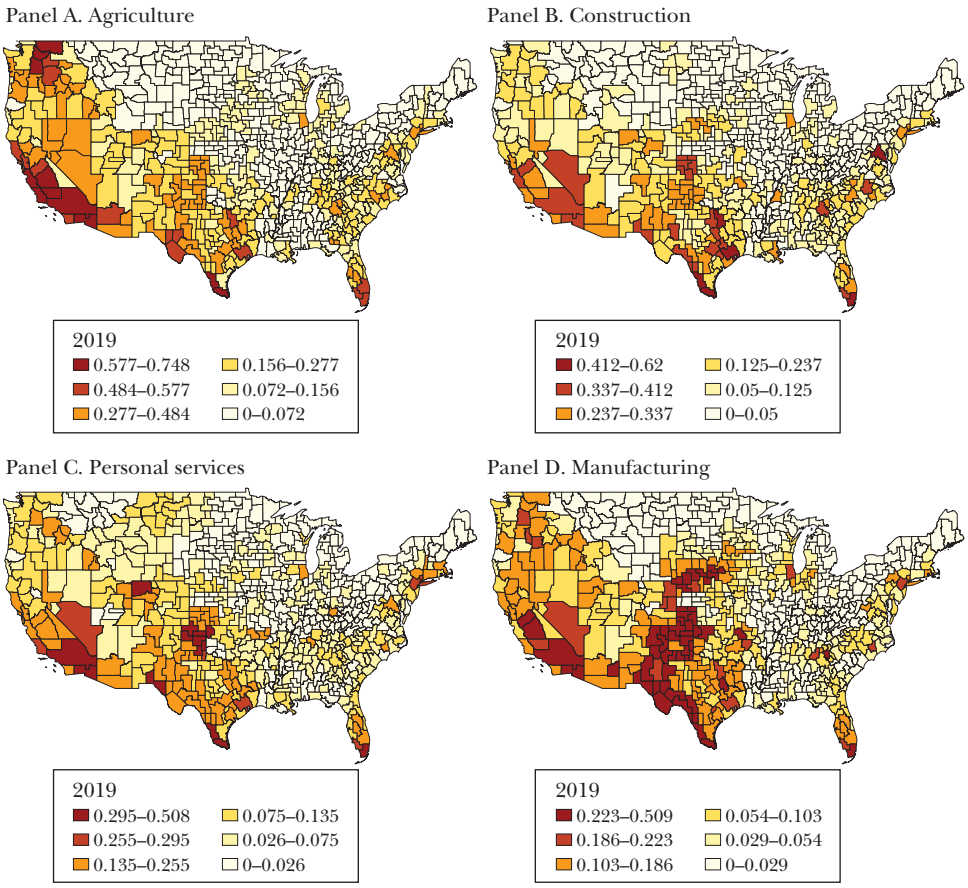
When immigration was expanding substantially, local and national labor markets adjusted along multiple margins. Now—the still unresolved COVID-19 pandemic changes in immigration notwithstanding—the United States may have begun a national experiment in how labor markets respond to substantial declines in the immigration of less-educated workers.

What Might the Future Hold for Latin American Immigration?

Although Hispanics remain the largest origin group of US immigrants, they may not be so within a few decades. If pre-COVID-19 immigration patterns were to persist, Latin America and the Caribbean would lose their current dominance in US labor inflows, just as the Irish, Germans, and Eastern Europeans did in previous eras. Under pre-pandemic trends, the Asian foreign-born share of the US population would surpass the Hispanic share by 2065 (Cohn 2015). In addition, Africa could become a more significant origin region for migration to the United States, given high population growth, low average incomes, and English fluency on much of the continent. Outside of Africa and the Middle East, population growth is on the decline, which may reduce origin-country demographic pressure for migration. Climate change, by disrupting production in many parts of the world and increasing the frequency of extreme weather, may become a more important migration push factor globally, although its specific impacts on US immigration are unclear.

Up to 2020, it seemed likely that most future US immigration inflows would be legal. The unauthorized labor inflows that so distinguished the Latin American

Figure 5
Share of Workers Born in Latin American and the Caribbean in Select Major Industries, 2019



Source: Based on IPUMS data on the 2019 1 percent sample of the American Community Survey.
Notes: Figures show the share of workers for a given industry and in a given commuting zone that was born in Latin America and the Caribbean. The legends divide shares into six categories by value for the bottom four quantiles and the top two deciles.

immigration surge had fallen dramatically. Visa over-stayers, who enter legally but become undocumented when their visas expire, had become more numerous than immigrants who entered illicitly (Meissner, Hipsman, and Aleinikoff 2018; Warren 2019). The drop-off in unauthorized border crossings was due in part to the fact that the US-Mexico border had become more heavily enforced than at any point in US history.

The pandemic-era increase in unauthorized border-crossing attempts has tested the new enforcement regime, with the outcome in terms of net US immigrant

flows still unknown. Title 42—the pandemic-induced US policy of no-consequence rapid expulsions of most migrants caught trying to cross the border—may have emboldened more migrants to attempt crossings and to keep attempting even if apprehended one or more times. With renewed high levels of attempted border crossings from an expanded set of origin countries, smuggling organizations have flourished, both along the US-Mexico border and along smuggling routes that extend deep into Latin America.

Perhaps the most notable pandemic-era change in US immigration from Latin America and the Caribbean is the exponential increase in the number of migrants asking for asylum. Asylum seekers can live and work legally in the United States while waiting for their claims to be adjudicated. If current backlogs persist, most migrants will not see their cases resolved for several years or more, possibly reducing their willingness to return to their home countries in the likely event, based on past precedent, that the large majority of claims are denied.

Given that the US economy is faced with an aging workforce and falling birth rates, pressures to liberalize US immigration policy may build, at least with respect to employment-based migration. Existing programs—such as those that allocate H-1B and H-2B temporary work visas—are already heavily oversubscribed and quickly run out of visas each year (Orrenius and Zavodny 2020). The pandemic has added urgency to immigration reform by creating labor shortages, albeit possibly temporary ones. Pandemic-based measures that closed US borders and consulates abroad in 2020 and 2021 prevented hundreds of thousands of immigrant workers from entering the country (Peri and Zaiour 2022), which may have further tightened US labor markets.

In response to these developments, labor markets in the United States and abroad will evolve. Difficulties in hiring native-born workers and obtaining visas for foreign-born workers may cause US labor costs to rise, which could induce firms to accelerate automation and the offshoring of production. Widespread experimentation with remote work during the pandemic may have taught firms that having all workers on-site is unnecessary. Such innovations may lead to more extensive changes in how foreign-born workers supply their services to US employers. At the same time, it is worthy of note that each new change in US immigration policy has inspired efforts to engineer around these changes in order to bring foreign-born labor into the country through other means. Economic and political crises abroad, fueled by continuing large international differences in living standards, are likely to sustain pressures for US immigrant inflows, whether from Latin America and the Caribbean or elsewhere.

References

Abramitzky, Ran, and Leah Boustan. 2017. "Immigration in American Economic History." *Journal of Economic Literature* 55 (4): 1311–45.

Abramitzky, Ran, Leah Boustan, and Katherine Eriksson. 2014. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy* 122 (3): 467–506.

Allen, Treb, Cauê de Castro Dobbin, and Melanie Morten. 2018. "Border Walls." NBER Working Paper 25267.

Allende, Isabel. 1999. *Daughter of Fortune*. London: Flamingo.

Alsan, Marcella, Katherine Eriksson, and Gregory Niemesh. 2020. "Understanding the Success of the Know-Nothing Party." NBER Working Paper 28078.

Angelucci, Manuela. 2012. "US Border Enforcement and the Net Flow of Mexican Illegal Migration." *Economic Development and Cultural Change* 60 (2): 311–57.

Bazzi, Samuel, Gordon Hanson, Sarah John, Bryan Roberts, and John Whitley. 2021. "Deterring Illegal Entry: Migrant Sanctions and Recidivism in Border Apprehensions." *American Economic Journal: Economic Policy* 13 (3): 1–27.

Bertoli, Simone. 2010. "Networks, Sorting and Self-Selection of Ecuadorian Migrants." *Annales d'Economie et de Statistique* 97/98: 261–88.

Borjas, George J. 2001. "Does Immigration Grease the Wheels of the Labor Market?" *Brookings Papers on Economic Activity* 31 (1): 69–119.

Burstein, Ariel, Gordon Hanson, Lin Tian, and Jonathan Vogel. 2020. "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence from the United States." *Econometrica* 88 (3): 1071–12.

Caballero, Maria Esther, Brian C. Cadena, and Brian K. Kovak. 2018. "Measuring Geographic Migration Patterns Using Matrículas Consulares." *Demography* 55 (3): 1119–45.

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

Calavita, Kitty. 2010. *Inside the State: The Bracero Program, Immigration, and the INS*. New Orleans: Quid Pro Books.

Card, David. 1990. "The Impact of the Mariel Boatlift on the Miami Labor Market." *ILR Review* 43 (2): 245–57.

Card, David, and Ethan G. Lewis. 2007. "The Diffusion of Mexican Immigrants during the 1990s." In *Mexican Immigration to the United States*, edited by George J. Borjas, 193–228. Chicago: NBER, University of Chicago Press.

Carrington, William J., Enrica Detragiache, and Tara Vishwanath. 1996. "Migration with Endogenous Moving Costs." *American Economic Review* 86 (4): 909–30.

Cavallo, Eduardo A. 2006. "Trade, Gravity and Sudden Stops: On How Commercial Trade Can Increase the Stability of Capital Flows." Research Department Publications 4491, Inter-American Development Bank, Research Department.

Champlin, Dell, and Eric Hake. 2006 "Immigration as Industrial Strategy in American Meatpacking." *Review of Political Economy* 18 (1): 49–70.

Chiquiar, Daniel, and Gordon H. Hanson. 2005. "International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States." *Journal of Political Economy* 113 (2): 239–81.

Clemens, Michael A. 2021. "Violence, Development, and Migration Waves: Evidence from Central American Child Migrant Apprehensions." *Journal of Urban Economics* 124: Article 103355.

Clemens, Michael A. 2022. "Migration on the Rise, a Paradigm in Decline: The Lasts Half-Century of Global Mobility." *AEA Papers and Proceedings*, 112: 257–61.

Clemens, Michael A., Claudio E. Montenegro, and Lant Pritchett. 2019. "The Place Premium: Bounding the Price Equivalent of Migration Barriers." *Review of Economics and Statistics* 101 (2): 201–13.

Cohn, D'Vera. 2015. "Future Immigration Will Change the Face of America by 2065." Pew Research Center, October 5. https://www.pewresearch.org/fact-tank/2015/10/05/future-immigration-will-change-the-face-of-america-by-2065/.

Cortes, Patricia. 2008. "The Effect of Low-Skilled Immigration on US Prices: Evidence from CPI Data." *Journal of Political Economy* 116 (3): 381–422.

Durand, Jorge, Douglas S. Massey, and Rene M. Zenteno. 2001. "Mexican Immigration to the United States: Continuities and Changes." *Latin American Research Review* 36 (1): 107–27.

Gathmann, Christina. 2008. "Effects of Enforcement on Illegal Markets: Evidence from Migrant

Smuggling along the Southwestern Border." *Journal of Public Economics* 92 (10-11): 1926–41.
Glaeser, Edward L. 2005. "Reinventing Boston: 1630–2003." *Journal of Economic Geography* 5 (2): 119–53.
Goldin, Claudia. 1994. "The Political Economy of Immigration Restriction in the United States, 1890 to 1921." In *The Regulated Economy: A Historical Approach to Political Economy*, edited by Claudia Goldin and Gary D. Libecap, 223–57. Chicago: NBER, University of Chicago Press.
Gonzalez-Barrera, Ana. 2017. "Mexican Lawful Immigrants among the Least Likely to Become U.S. Citizens." Pew Research Center, June 29. https://www.pewresearch.org/hispanic/2017/06/29/mexican-lawful-immigrants-among-least-likely-to-become-u-s-citizens/.
Gramlich, John. 2022. "Key Facts about Title 42, the Pandemic Policy That Has Reshaped Immigration Enforcement at US-Mexico Border." Pew Research Center, April 27. https://www.pewresearch.org/fact-tank/2022/04/27/key-facts-about-title-42-the-pandemic-policy-that-has-reshaped-immigration-enforcement-at-u-s-mexico-border/.
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.
Hanson, Gordon. 2005. *Why Does Immigration Divide America?* Washington, DC: Peterson Institute.
Hanson, Gordon, and Craig McIntosh. 2009. "The Demography of Mexican Migration to the United States." *AER: Papers & Proceedings* 99 (2): 22–7.
Hanson, Gordon, and Craig McIntosh. 2010. "The Great Mexican Emigration." *Review of Economics and Statistics* 92 (4): 798–810.
Hanson, Gordon, and Craig McIntosh. 2012. "Birth Rates and Border Crossings: Latin American Migration to the US, Canada, Spain and the UK." *Economic Journal* 122 (561): 707–26.
Hanson, Gordon, and Craig McIntosh. 2016. "Is the Mediterranean the New Rio Grande? US and EU Immigration Pressures in the Long Run." *Journal of Economic Perspectives* 30 (4): 57–82.
Hanson, Gordon, Chen Liu, Craig McIntosh. 2017. "The Rise and Fall of U.S. Low-Skilled Immigration." *Brookings Papers on Economic Activity* 47 (1): 83–168.
Hanson, Gordon, Pia Orrenius, and Madeline Zavodny. 2023. "Replication data for: US Immigration from Latin America in Historical Perspective." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E183147V1.
Hanson, Gordon H., and Antonio Spilimbergo. 1999. "Illegal Immigration, Border Enforcement, and Relative Wages: Evidence from Apprehensions at the US-Mexico Border." *American Economic Review* 89 (5): 1337–57.
Higham, John. 2002. *Strangers in the Land: Patterns of American Nativism, 1860–1925.* New Brunswick: Rutgers University Press.
Huntington, Samuel P., and Steve Dunn. 2004. *Who Are We? The Challenges to America's National Identity.* New York: Simon and Schuster.
Jasso, Guillermina, Douglas S. Massey, Mark R. Rosenzweig, and James P. Smith. 2008. "From Illegal to Legal: Estimating Previous Illegal Experience among New Legal Immigrants to the United States." *International Migration Review* 42 (4): 803–43.
Lessem, Rebecca. 2018. "Mexico–U.S. Immigration: Effects of Wages and Border Enforcement." *Review of Economic Studies* 85 (4): 2353–88.
Lewis, Ethan. 2011. "Immigration, Skill Mix, and Capital Skill Complementarity." *Quarterly Journal of Economics* 126 (2): 1029–69.
Mahajan, Parag, and Dean Yang. 2020. "Taken by Storm: Hurricanes, Migrant Networks, and US Immigration." *American Economic Journal: Applied Economics* 12 (2): 250–77.
Marcelli, Enrico A., and Wayne A. Cornelius. 2001. "The Changing Profile of Mexican Migrants to the United States: New Evidence from California and Mexico." *Latin American Research Review* 36 (3): 105–31.
Massey, Douglas S., Jorge Durand, and Nolan J. Malone. 2003. *Beyond Smoke and Mirrors: Mexican Immigration in an Era of Economic Integration.* New York: Russell Sage Foundation.
Massey, Douglas S., and Karen A. Pren. 2012. "Unintended Consequences of US Immigration Policy: Explaining the Post-1965 Surge from Latin America." *Population and Development Review* 38 (1): 1–29.
McKenzie, David, and Hillel Rapoport. 2007. "Network Effects and the Dynamics of Migration and Inequality: Theory and Evidence from Mexico." *Journal of Development Economics* 84 (1): 1–24.
Meissner, Doris, Faye Hipsman, and T. Alexander Aleinikoff. 2018. "The U.S. Asylum System in Crisis." Migration Policy Institute. https://www.migrationpolicy.org/sites/default/files/publications/MPI-AsylumSystemInCrisis-Final.pdf.

Monras, Joan. 2020. "Immigration and Wage Dynamics: Evidence from the Mexican Peso Crisis." *Journal of Political Economy* 128 (8): 3017–89.

Munshi, Kaivan. 2003. "Networks in the Modern Economy: Mexican Migrants in the U.S. Labor Market." *Quarterly Journal of Economics* 118 (2): 549–99.

Orrenius, Pia M., and Madeline Zavodny. 2003. "Do Amnesty Programs Reduce Undocumented Immigration? Evidence from IRCA." *Demography* 40 (3): 437–50.

Orrenius, Pia M., and Madeline Zavodny. 2005. "Self-Selection among Undocumented Immigrants from Mexico." *Journal of Development Economics* 78 (1): 215–40.

Orrenius, Pia M., and Madeline Zavodny. 2009. "Do Immigrants Work in Riskier Jobs?" *Demography* 46 (3): 535–51.

Orrenius, Pia M., and Madeline Zavodny. 2010. *Beside the Golden Door: U.S. Immigration Reform in a New Era of Globalization*. Washington, DC: AEI Press.

Orrenius, Pia M., and Madeline Zavodny. 2020. "Help Wanted: Employer Demand for Less-Skilled Temporary Foreign Worker Visas in an Era of Declining Unauthorized Immigration." *Russell Sage Foundation Journal of the Social Sciences* 6 (3): 45–67.

Orozco-Aleman, Sandra, and Heriberto Gonzalez-Lozano. 2018. "Drug Violence and Migration Flows: Lessons from the Mexican Drug War." *Journal of Human Resources* 53 (3): 717–49.

Passel, Jeffrey S., and D'Vera Cohn. 2019. "Mexicans Decline to Less than Half the U.S. Unauthorized Immigrant Population for the First Time." Pew Research Center, June 12. https://www.pewresearch.org/fact-tank/2019/06/12/us-unauthorized-immigrant-population-2017/.

Peri, Giovanni, and Reem Zaiour. 2022. "Labor Shortages and the Immigration Shortfall." UC Davis. https://econofact.org/labor-shortages-and-the-immigration-shortfall.

Roberts, Bryan, Edward Alden, and John Whitley. 2013. *Managing Illegal Immigration to the United States: How Effective Is Enforcement?* New York: Council on Foreign Relations.

Warren, Robert. 2019. "Overstays Exceeded Illegal Border Crossers after 2010 Because Illegal Entries Dropped to Their Lowest Level in Decades." Center for Migration Studies. https://cmsny.org/publications/essay-warren-042419/.

Wolfe, George. 2021. "Where Are Venezuelan Migrants and Refugees Going? An Analysis of Legal and Social Contexts in Receiving Countries." Center for Migration Studies. https://cmsny.org/publications/venezuelan-migrants-legal-contexts-wolfe-010421/.

US Department of Homeland Security. 2022. *2020 Yearbook of Immigration Statistics*. Washington, DC: Office of Immigration Statistics.

Zimran, Ariell. 2021. "Immigrant Distribution in the United States during the Age of Mass Migration." NBER Working Paper 28812.x

Journal of Economic Perspectives—Volume 37, Number 1—Winter 2023—Pages 223–240

Oleg Itskhoki: 2022 John Bates Clark Medalist

Andrew Atkeson and Gita Gopinath

The 2022 John Bates Clark Medal of the American Economic Association was awarded to Oleg Itskhoki, Professor of Economics at the University of California, Los Angeles, for his fundamental contributions to international macroeconomics and international trade. Since the end of World War II, the world economy has been engaged in an ongoing, if at times fitful, process of opening to international trade in goods and flows of private capital. The field of international macroeconomics explores how this process of globalization affects the choices nations can and should make regarding their monetary and fiscal policies. Oleg's insights into trade and exchange rate behavior have far-reaching implications that will be the focus of considerable research in the years to come.

Oleg was born in Russia at the end of the Soviet Union. He reports that he played a lot of tennis until the age of 16 and did not really think of an academic career until he had to apply to college at the end of high school. He was initially drawn to economics in the hope that he might find a good job in the post-Soviet economy. It was not until after his undergraduate studies at Moscow State University that he encountered modern economics in the master's program at the New Economic School in Moscow. There he was introduced to the Dixit and Norman (1980) and Obstfeld and Rogoff (1996) textbooks, which sparked his fascination with international economics. He attended Harvard as a PhD student, and he points to a long list of faculty there as key influences, including John Campbell, Ken Rogoff,

■ *Andrew Atkeson is the Stanley M. Zimmerman Professor of Economics, University of California, Los Angeles, Los Angeles, California. Gita Gopinath is First Deputy Managing Director, International Monetary Fund, Washington, DC.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.1.223.

Oleg Itskhoki

and Jim Stock, as well as Daron Acemoglu and Jordi Galí at the Massachusetts Institute of Technology. His participation as a student in a reading group organized by newly arrived young Harvard faculty members Aleh Tsyvinski, Gita Gopinath, Pol Antràs, and Manuel Amador was particularly meaningful in showing him how new ideas are born and developed. He counts himself most fortunate to have had the opportunity early on in his PhD studies to work with Elhanan Helpman as well. After spending several years at Princeton University, along with visiting appointments at the University of Chicago and Stanford, Oleg moved to the University of California, Los Angeles, in 2019, where he holds the Venu and Ana Kotamraju Endowed Chair in Economics.

In this paper, we aim to put Oleg's research in the broader context of some of the main questions and puzzles that have confronted international macroeconomics since the breakdown of the Bretton Woods system of fixed exchange rates more than 50 years ago. We examine some of Oleg's most prominent work on these puzzles with a focus on four areas: (1) firms' strategies for pricing their products in international markets in the face of volatile nominal exchange rates; (2) how firms' choices to set prices in a dominant currency such as the US dollar change the impact of exchange rate shocks on the macroeconomy; (3) policy options to gain macroeconomic flexibility for countries that fix their exchange rate or adopt a common currency; and (4) a proposed unified resolution of major puzzles regarding the interaction of exchange rates and the macroeconomy that have stymied international macroeconomics for many years. We also describe Oleg's contributions to our understanding of the impact on inequality of increasing international trade. We refer to his key papers throughout by number, as listed in Table 1.

Table 1
Selected Papers by Oleg Itskhoki

1	Gita Gopinath, Oleg Itskhoki, and Roberto Rigobon. 2010. "Currency Choice and Exchange Rate Pass-Through." *American Economic Review* 100 (1): 304–36.
2	Gita Gopinath and Oleg Itskhoki. 2010. "Frequency of Price Adjustment and Pass-Through." *Quarterly Journal of Economic*s 125 (2): 675–727.
3	Gita Gopinath and Oleg Itskhoki. 2011. "In Search of Real Rigidities." In *NBER Macroeconomics Annual 2010*, vol. 25, edited by D. Acemoglu and M. Woodford, 261–309. Chicago: University of Chicago Press.
4	Mary Amiti, Oleg Itskhoki, and Jozef Konings. 2014. "Importers, Exporters, and Exchange Rate Disconnect." *American Economic Review* 104 (7): 1942–978.
5	Mary Amiti, Oleg Itskhoki, and Jozef Konings. 2019. "International Shocks, Variable Markups, and Domestic Prices." *Review of Economic Studies* 86 (6): 2356–402.
6	Gita Gopinath and Oleg Itskhoki. 2021. "Dominant Currency Paradigm: A Review." Forthcoming in *Handbook of International Economics*, vol. 6.
7	Mary Amiti, Oleg Itskhoki, and Jozef Konings. 2022. "Dominant Currencies: How Firms Choose Currency Invoicing and Why It Matters." *Quarterly Journal of Economics* 137 (3): 1435–493.
8	Emmanuel Farhi, Gita Gopinath, and Oleg Itskhoki. 2014. "Fiscal Devaluations." *Review of Economic Studies* 81 (2): 725–60.
9	Omar Barbiero, Emmanuel Farhi, Gita Gopinath, and Oleg Itskhoki. 2019. "The Macroeconomics of Border Taxes." In *NBER Macroeconomics Annual 2018*, vol. 33, edited by Jonathan Parker and Martin S. Eichenbaum, 395–457. Chicago: University of Chicago Press.
10	Oleg Itskhoki and Dmitry Mukhin. 2021. "Exchange Rate Disconnect in General Equilibrium." *Journal of Political Economy* 129 (8): 2183–232. Lead article.
11	Oleg Itskhoki and Dmitry Mukhin. 2021. "Mussa Puzzle Redux." NBER Working Paper 28950, National Bureau of Economic Research, Cambridge, MA. MFA Best Paper Award in Asset Pricing, 2021.
12	Elhanan Helpman, Oleg Itskhoki, Marc-Andreas Muendler, and Stephen J. Redding. 2017. "Trade and Inequality: From Theory to Estimation." *Review of Economic Studies* 84 (1): 357–405.
13	Elhanan Helpman, Oleg Itskhoki, and Stephen J. Redding. 2010. "Inequality and Unemployment in a Global Economy." *Econometrica* 78 (4): 1239–283.
14	Elhanan Helpman and Oleg Itskhoki. 2010. "Labour Market Rigidities, Trade and Unemployment." *Review of Economic Studies* 77 (3): 1100–137.

We begin with a review of the state of the literature in international economics prior to Oleg's work to provide a context for understanding Oleg's contributions to this field.

Background

The experience of the world economy during the past 50 years following the breakdown of the Bretton Woods system of fixed exchange rates has raised many questions and puzzles. A useful point of entry to these issues is the concept of

international real relative prices. At the macroeconomic level, changes in international real relative prices are measured by the changes in the *real exchange rate* and the *terms of trade*.

Changes in the real exchange rate are defined as changes in the relative price of consumption baskets in various countries measured in a common currency; that is, as the sum of inflation differentials across countries and changes in the nominal exchange rate. Thus, mechanically, changes in real exchange rates are related to changes in nominal exchange rates to the extent that inflation differentials across countries move or do not move to offset changes in nominal exchange rates.

Changes in the terms of trade are defined as changes in the ratio of an index of prices for goods that are exported, to an index of prices for goods that are imported (expressed in a common currency) from the perspective of a single country. At the micro level, fine-grained data on international real relative prices are increasingly available from price quotes for individual goods at the retail and the wholesale levels, and at the border as goods are exported and imported.

Over recent decades, many countries have moved away from monetary policies aimed at maintaining a fixed or stable nominal exchange rate against other currencies to monetary policies that focus on domestic inflation and business cycles, while allowing the nominal exchange rate to vary widely over time. These policy shifts have been accompanied by five notable shifts in the patterns of international real relative prices.

1. International real relative prices are much more volatile than they were under fixed exchange rates.

2. These large swings in international real relative prices are very persistent over time.

3. These large and persistent swings in international relative prices are apparent even in fine-grained micro data for traded goods.

4. The direction of movements of these international real relative prices does not appear to be closely connected to movements in other macroeconomic variables.

5. It appears that a country can significantly reduce the volatility of its real exchange rate over the medium term by choosing a monetary policy aimed at stabilizing its nominal exchange rate against other currencies.

At some level, the first observation—that international real relative prices move in the short term as the nominal exchange rate moves—is not much of a puzzle. Since the foundational work of Robert Mundell and Marcus Fleming in the 1960s,[1] the idea that movements in nominal exchange rates affect real relative prices in the short term because nominal prices are "sticky" has been central to many models used in international macroeconomics. To put it another way, broad inflation differentials across countries are typically slow to move in the short term, whereas nominal exchange rates can be quite volatile over short time periods.

[1] See the description of Robert Mundell's contributions by the Nobel Prize Committee at https://www.nobelprize.org/uploads/2018/06/advanced-economicsciences1999.pdf and Rose (2000). For modern development of these ideas, see Obstfeld and Rogoff (1996).

However, the second and third observations were a real surprise to academics and policymakers alike. Before the move to floating exchange rates among most major currencies, many expected that—at least over the medium and long terms—arbitrage in goods markets would anchor international real relative prices independent of the behavior of nominal exchange rates. The logic behind this expectation is straightforward: through international trade, goods should flow from countries where they are cheap to countries where they are expensive until real relative prices of goods (measured in units of a common currency) are stabilized. This arbitrage in goods markets should stabilize real relative prices across countries regardless of whether nominal exchange rates are volatile or stable. This hypothesis is known as the hypothesis of "purchasing power parity."

But by the mid-1990s, this initial expectation had clearly been proven wrong, as major currencies—such as the US dollar and the Japanese yen—experienced wide swings in their real exchange rates over periods of five years or more in the 1980s and 1990s, which in turn led to large and persistent movements in international real relative prices. This apparent failure of arbitrage in goods markets to limit fluctuations in international real relative prices over long time horizons came to be known as the "purchasing power parity puzzle" (Rogoff 1996).

The purchasing power parity puzzle deepens with a dive into the micro or disaggregated price data. In standard modeling frameworks, movements in the real exchange rate correspond to changes in the relative price of traded and nontraded goods across countries. The logic is that arbitrage in goods markets should stabilize the real relative price of *traded* goods across countries, but should have less impact on the real relative prices of goods and services that are not traded. However, Engel (1999) showed that these large and persistent fluctuations in the real exchange rate did not result simply from changes in the relative prices of goods that are not traded across countries. Instead, fluctuations in the real exchange rate measured using traded goods account for nearly all the observed fluctuations in real exchange rates.

Moreover, Engel and Rogers (1996) used micro data on consumer prices for various cities in the United States and Canada during a period in which the US dollar–Canadian dollar nominal exchange rate fluctuated in a wide range. They showed that real relative price volatility at the micro level across cities on either side of the US and Canadian border was much larger than real relative price volatility across cities within the two countries, where the nominal exchange rate is fixed because of the use of a common currency. Thus, it appeared that nominal exchange rate variability between US and Canadian dollars had a substantial impact on the volatility of real relative consumer prices across cities over and above what the geographic distance between those cities might predict. Gopinath et al. (2011) and Cavallo, Neiman, and Rigobon (2014) found similar results using even more micro retail price observations—data from retailers selling identical products in many countries. Cavallo, Neiman, and Rigobon (2014) demonstrate that products within the set of countries whose common currency is the euro frequently sell at a real relative price of one, as predicted by purchasing power parity, while this real relative price fluctuates substantially with changes in nominal exchange rates

between countries without a common currency. These studies, and much related work, raised questions about why firms operating across international boundaries would choose to allow the real relative price of their products sold across different locations to vary so much with changes in the nominal exchange rate.

The fourth observation presented an additional puzzle regarding the behavior of exchange rates. That real exchange rates among major currencies undergo wide swings over five-year horizons or more would not be as puzzling if it were possible to account for these movements in international real relative prices, even after the fact, based on movements in observed macroeconomic fundamentals. But after 50 years of searching for a robust connection between exchange rate movements and movements in macroeconomic variables, we have come up mostly empty-handed. Meese and Rogoff (1983) demonstrated this apparent disconnect between exchange rates and macroeconomic fundamentals very soon after the breakdown of the Bretton Woods system of fixed exchange rates. The "exchange rate disconnect puzzle" they identified persists to this day.[2]

The large and persistent swings in real exchange rates between major countries also lead to questions about whether and how they can be avoided—or at least moderated—by appropriate policies. Here, our fifth observation about policies suggests intriguing possibilities. Empirical research by Mussa (1986) and Flood and Rose (1995), together with the micro data on retail prices cited above, raises the possibility that countries can limit the volatility of their international real relative prices over a medium- to long-term horizon through monetary policy aimed at maintaining a fixed nominal exchange rate. These studies do not establish a causal link between the choice of nominal exchange rate policy and the medium- and long-term volatility of real exchange rates. However, they document a robust, broad historical connection between nominal exchange rate volatility and real exchange rate volatility—along with remarkably little connection between changes in the volatility of other macroeconomic fundamentals when a country shifts between a fixed and floating nominal exchange rate regime. This observation is often called the "Mussa puzzle."

Oleg, in work with a range of coauthors, has made important contributions to addressing each of these puzzles, with his most recent work in papers [10] and [11] being the most ambitious in seeking to account for all five of these puzzles in a unified framework. We discuss four strands of his work on these puzzles in turn.

Micro Data on Firms' Pricing Policies

This first strand of Oleg's research develops models of firms' strategies for pricing their products in international markets in the face of volatile nominal exchange rates. He also evaluates those models with increasingly rich micro data.

[2] See Obstfeld and Rogoff (2000) and the published comments on this paper by Jeanne (2000) and Engel (2000) for a good summary of the state of the earlier literature on these puzzles.

In a pair of papers written in collaboration with Gita Gopinath [1, 2] and Roberto Rigobon [1], Oleg and his coauthors use detailed micro data on prices, collected by the US Bureau of Labor Statistics for its construction of price indices for exported and imported goods, to shed new light on the decisions of firms actively engaged in international trade to set prices paid at the border for imported and exported goods. These micro data on prices of exported and imported goods, first explored in Gopinath and Rigobon (2008), are unlike prior work with micro data based on retail or wholesale prices. These new micro data allow researchers to see the links between changes in exchange rates and the pricing of traded goods, free of the nontraded local distribution costs that contaminate the link between the prices of traded goods themselves and what consumers eventually pay at the retail level. In addition, researchers can see not only the extent and duration of stickiness in the prices of traded goods but also how firms respond to shocks when they choose to reset those prices.

Several important empirical regularities emerged from the study of these micro data. First, and most basically, these data confirmed that prices for imported and exported goods are typically sticky. Thus, there is a mechanical link (at least in the short term) between changes in the nominal exchange rate and the real relative prices of imported and exported goods, or the terms of trade, as is central to the analytical framework pioneered in the 1960s by Mundell (1963).

The nature of this mechanical link between changes in the nominal exchange rate and the terms of trade and sticky prices depends on the currencies in which firms set their nominal prices. For example, say that firms set their nominal prices for traded goods in the currency of the country where the good is produced, in what is called "producer currency pricing." Then, exported goods' prices are sticky in the exporter's currency and imported goods' prices are sticky in the currency of the countries exporting these goods. If the nominal exchange rate changes so that a country's currency becomes more valuable than those of its trading partners, the real prices of its exports rise relative to those of its imports. In contrast, if firms set their nominal prices in the currency of the country to which the good is shipped, in what is called "local currency pricing," then the reverse is true, given the same change in the nominal exchange rate. Therefore, to figure out how changes in nominal exchange rates affect a country's terms of trade, it is critical to understand how firms make decisions about the currency in which to invoice their products.

It is here that papers [1] and [2] make substantial contributions. In these micro data, it is evident that firms do not fully adjust their nominal prices in response to nominal exchange rate changes, even when they choose to change their nominal prices. More importantly, Oleg, Gita, and Roberto find systematic links between the choices firms made to invoice their products in different currencies and the choices they made to change their nominal prices in response to exchange rate changes. In [1], focusing on the response of US import prices to changes in the nominal exchange rate, Oleg and his coauthors document a systematic difference in the response of the US dollar price of the goods invoiced

in dollars (as is true for local currency pricing) and those that are invoiced in a foreign currency (as is true for producer currency pricing), even when these nominal prices are reset. In [2], Oleg and Gita show that firms that change their prices more often also make bigger changes in their nominal prices in response to nominal exchange rate changes over the long term.

Both of these findings call for theories that jointly explain firms' choice of invoicing currency when their prices are sticky and their decisions to change their nominal prices in response to changes in nominal exchange rates. These papers demonstrate that firms choose to price in a currency in which their desired prices are stable.

Papers [1] and [2] highlight two key mechanisms that influence pricing. The first is the *imported intermediate input channel*. An exporting firm that relies on imported inputs priced, say, in US dollars has a marginal cost that is relatively stable in dollars and consequently will price its exports in US dollars, because the sticky dollar price is close to optimal even during periods of nonadjustment.

The second mechanism concerns *strategic complementarities in firms' price-setting decisions,* which refers to the extent to which firms' desired prices depend on their own marginal cost of production as well as on the prices other firms are charging. Standard models in which firms are perfectly competitive in product markets or face a constant elasticity of their residual demand curve, independent of the prices chosen by their competitors, have no strategic complementarities. In those models, firms choose a price equal either to their marginal cost or to that marginal cost times a constant proportional markup, regardless of the prices charged by competitors. In contrast, when strategic complementarities are strong, a firm that finds its marginal costs affected by a change in the nominal exchange rate will often not fully pass on the change in marginal cost to its customers but will also adjust the markup of its price over marginal cost. This choice arises from concern about competition with firms whose input prices and marginal costs are not affected by exchange rate changes.

Strategic complementarities explain why, even conditional on changing prices, a firm does not alter its US dollar price by much. In prior work, Goldberg and Hellerstein (2008; 2013) developed structural models of the impact of strategic complementarities on firms' decisions to reset prices in response to changes in exchange rates, with a focus usually on a specific industry. Where Oleg and his coauthors extended this literature is in considering the interaction of imperfect competition and exchange rate variability both on firms' pricing decisions and their decisions to invoice their goods in a particular currency. They show, for example, that a firm whose competitors price in US dollars is motivated to price in dollars as well, so that exchange rate movements do not lead to relative price adjustments that cause the firm to lose market share.

Oleg develops this agenda further in joint work with Mary Amiti and Jozef Konings in [4] and [5], where he brings to bear novel micro data on the pricing decisions of Belgian firms and developed structural frameworks to analyze the strength of the intermediate input channel, of strategic complementarities, and

of the interaction of these forces in shaping firms' invoicing and pricing decisions. This richer data set includes information on the extent to which firms exporting from Belgium import the inputs they use in production, as well as information about these firms' marginal costs and the prices charged by their competitors. With these data, Oleg and his coauthors are able in paper [4] to provide compelling evidence of the two theoretical mechanisms discussed previously and to model the role of these forces in shaping firms' pricing and invoicing decisions, using a model of firms' pricing and choice of currency of invoicing under imperfect competition based on that in Atkeson and Burstein (2008). In [5], they construct augmented micro data not only on firms' marginal costs but also on the prices of their competitors, developing a theoretical framework to directly decompose firms' price changes into a response to changes in its own marginal cost and a response to changes in the prices charged by its competing firms.

The role of strategic complementarities in the pricing decisions of firms is central not only to international macroeconomics, but also to core questions in closed-economy monetary macroeconomics. In both fields, a key question is how monetary or nominal shocks can have a persistent real effect well beyond the horizon for which firms' nominal prices are sticky. In [3], Oleg and Gita compare the evidence and analytical frameworks used in international and closed economy macroeconomics to understand this persistence. Regarding strategic complementarities, studies using data on firms' international pricing decisions have the advantage that nominal exchange rate shocks are frequent, large, and persistent. These studies typically find strong evidence of strategic complementarities. In contrast, with data from closed economies, there are fewer well-identified nominal shocks, and these shocks tend to be smaller and less persistent. Thus, studies relying on data from a single country tend to find only weak evidence of strategic complementarities. In this dimension, work in international economics may inform our future models of the impact of nominal shocks in closed economies.

In these papers, Oleg and his coauthors provide definitive empirical evidence and provocative modeling frameworks to help us understand the economics underlying the purchasing power parity puzzle, the exchange rate disconnect puzzle, and the behavior of the terms of trade. Certainly, a significant portion of the resolution of the purchasing power parity puzzle stems from the fact that many goods and services are not traded internationally, and thus changes in the nominal exchange rate do not significantly affect the pricing decisions of firms producing these goods and services. The direct micro evidence marshaled in these papers, however, indicates that the industrial organization of the markets in which firms that export and import traded goods also plays an important role in resolving these puzzles. One characteristic of this industrial organization is that firms that can choose currencies in which they price their products based on the characteristics of the specific shocks and competition they face. Moreover, due to heterogeneous use of imported intermediate inputs and heterogeneous product market competition, these firms, in equilibrium, do not choose to fully change their nominal prices (in the currency

in which they are invoiced) in response to changes in nominal exchange rates, even over relatively long time horizons.

Dominant Currency Pricing

In standard open economy macroeconomic models, when the nominal prices of exported and imported goods are sticky, a change in the nominal exchange rate can mechanically alter the real relative price of a country's exports and imports and thus alter world consumers' desired allocation of spending across countries. Prior to the work discussed above with micro data, the magnitude and direction of this effect of nominal exchange rate changes on a country's terms of trade were not clear. As previously discussed, when prices for imported and exported goods are sticky, the mechanical impact of a change in the nominal exchange rate on a country's terms of trade depends on the currencies in which firms choose to price exported and imported goods. Under producer currency pricing, the effect of the exchange rate on the terms of trade goes one way; under local currency pricing, it goes the other way.

But Gopinath (2015) and Gopinath et al. (2020) document the absence of *both* of these pricing paradigms in the micro data. Instead, most firms that are engaged in trade worldwide price their goods in one of a few dominant currencies—primarily the US dollar or the euro. In [6], Oleg and Gita survey this evidence. The terms of trade for many countries tend to be stable despite large nominal exchange rate changes, contrary to the earlier classic models of Mundell and Fleming. Further, when a country's exchange rate depreciates, there is a relatively muted impact on its exports in the short term; mainly, the country's imports decline as the relative prices of imports rise relative to domestic goods.

In [7], Oleg, working again with Mary Amiti and Jozef Konings, tackles the question of why firms engaged in international trade would choose to invoice their products in a dominant currency. They use micro data with evidence on the choice of currency invoicing at the firm-product-destination-month level. They show that firms' choice of currency for invoicing their products is an active choice that persists over time and that this choice is more closely tied to firm and destination-country characteristics than to industry or product characteristics. They show that for Belgian imports and exports outside the euro area, dominant currency pricing is widespread: the vast majority of these exports and imports outside the euro area are invoiced in either euros or US dollars. They extend their previous modeling of firms' currency invoicing and pricing decisions to allow for dominant currency pricing and show that firms' product invoicing decisions are systematically related to attributes such as firm size (a proxy for market share), firms' share of imported intermediate inputs and the currency invoicing of those intermediate inputs, and the currency invoicing decisions of competitors. Based on this match between theory and data, they argue that

strategic complementarities and imported intermediate inputs in firms' currency invoicing decisions can entrench an invoicing currency in a dominant role for a long time.

Fiscal Policy as a Substitute for Exchange Rate Devaluations

One classic dilemma for policymakers is whether to pursue a fixed nominal exchange rate (or even to adopt a common currency) or to allow the exchange rate to float. In the standard framework for analyzing the cost and benefits of alternative exchange rate regimes, a fixed exchange rate regime is seen as having the benefit of reducing the volatility of international real relative prices, and the use of a common currency, such as the euro, is seen as facilitating further economic integration across the boundaries of countries that adopt such a common currency. However, these benefits are considered to accompany the cost of less policy flexibility. In particular, if a country with a fixed exchange rate or in a common currency area experiences an economic downturn, it is typically seen as not having the option of changing its nominal exchange rate to alter its terms of trade and thus shift worldwide expenditures toward its national economy. In the standard framework for analyzing this policy dilemma, pioneered by Robert Mundell (1961), the question of whether a country should have a fixed exchange rate or adopt a common currency depended in part on the extent to which that country had access to policy tools other than changes in its nominal exchange rate to deal with negative macroeconomic shocks.

In [8], Oleg, with Emmanuel Farhi and Gita Gopinath, points out that this conventional wisdom overlooks the fact that a country that has a common currency with its neighbors can achieve the effects of an exchange rate devaluation on its terms of trade with a small set of changes in fiscal policies—either a coordinated change in import tariffs and export subsidies or a change in value-added taxes and payroll tax deductions. In the debate leading up to the US corporate tax reform in 2017, this question of the impact of changes in tax policy on the US terms of trade and macroeconomic outcomes took on added urgency as Republicans in the US House of Representatives proposed border adjustments of the tax on corporate profits that would tax imports and allow firms to deduct taxes on exports. In [9], these same three authors, joined by Omar Barbiero, offer an analysis of such a border adjustment of the corporate profits tax. This analysis significantly extends prior work on this question by Lerner (1936), Grossman (1980), and Feldstein and Krugman (1990) by examining the impact of the imposition of a border adjustment to corporate profit taxes in a fully dynamic sticky-price and sticky-wage New Keynesian model with alternative assumptions about the invoicing currency of traded goods. The key finding in this paper is that the short-term macroeconomic impact of such a fiscal policy can be substantial—the magnitude depends on how much firms change their prices in response to the changes in taxes and the induced changes in the US dollar exchange rate. Thus, in this paper, Oleg and his coauthors

draw out the implications of their work on firms' pricing decisions for significant questions in public finance.

Exchange Rate Disconnect

In the strands of research already discussed, Oleg focused on individual pieces of the puzzle of international real relative prices, taking shocks to nominal exchange rates as given. In [10] and [11], Oleg and Dmitry Mukhin aim to assemble the pieces of the puzzle into a coherent whole. They seek an underlying explanation of the shocks that drive changes in nominal exchange rates, together with an account of the connection between these shocks and macroeconomic fluctuations. Their challenge is to do so in a manner that can be reconciled with the five puzzles listed at the start of our background section on the state of the literature in international finance prior to Oleg's work and that can also provide a framework for understanding how adopting a monetary policy aimed at fixing the nominal exchange rate can bring stability to a country's international real relative prices.

Oleg's work in this area starts from the observation that in standard models of the international macroeconomy, real exchange rate volatility is intimately linked to macroeconomic volatility—regardless of whether that volatility is driven by monetary shocks or productivity shocks. This link between the macroeconomy and the real exchange rate is most prominent in versions of these standard models that have complete asset markets, in the sense that macroeconomic risks are optimally shared across consumers in different countries. Such models make a stark prediction that changes in the real exchange rate are directly linked to changes in the ratio of the marginal utility of consumption for domestic and foreign consumers, regardless of the nature of shocks to domestic and foreign economies. The failure of this implication of optimal risk-sharing to hold in the data was first documented in Backus and Smith (1993) and is now referred to as the Backus-Smith puzzle. Of course, the assumption of complete international asset markets is extreme. But research over several decades has confirmed that this tight link between macroeconomic volatility and real exchange rate volatility continues to hold in standard models with quite limited opportunities for cross-border risk sharing as long as the macroeconomic volatility is driven by standard monetary or productivity shocks (for example, see Lustig and Verdelhan 2019).

Given these observations, the literature in international macroeconomics has begun to consider an alternative source of shocks to real exchange rates—shocks to the desired allocation of portfolios across countries. These shocks in asset markets have intellectual antecedents in what were termed "portfolio balance models of exchange rates," developed initially in the 1970s (for example, Kouri 1976). In the popular press, such shocks are typically referred to as shocks to "the demand for dollars" or as "flight" by international investors from or to a particular currency.

Research into such shocks as a source of exchange rate volatility was reinvigorated in work by Jeanne and Rose (2002) and Gabaix and Maggiori (2015). This

work was motivated both by the observation that a small number of major banks worldwide intermediate the vast majority of trading of nominal exchange rate risk in spot and derivatives markets and by the puzzling behavior of interest rates in different currencies and exchange rates observed over the past 50 years. Specifically, if the nominal interest rate for one currency, say the US dollar, is high relative to that for another currency, say the Japanese yen, it might be expected that the value of the dollar would fall relative to that of the yen over time and would equalize the expected returns to investing at these two interest rates when the returns are expressed in a common currency. This prediction is sharply contradicted by the data. Instead, there are large and persistent movements in the expected excess return to investing in one currency versus another; it is typically profitable, at least for major currencies, to invest in a currency when it has a high nominal interest rate relative to others. This behavior of interest rates and changes in nominal exchange rates is referred to as the "Fama puzzle" (Fama 1984). Although there has been extensive effort to understand the Fama puzzle in the context of frictionless capital markets based on variation over time in currency risk premia, recent work has focused on the resolution of the puzzle as arising from frictions in international capital markets that offer profitable trading opportunities to major banks that trade exchange rate risk.

The models of how shocks to the desired allocation of portfolios across countries affect exchange rates start from the hypothesis that frictions in financial markets cause most investors to focus on holding assets denominated in the currency of the country where they live and do business. Such models assume that investors do not actively participate in trading nominal exchange rate risk. Hence, international capital markets have relatively few investors willing to absorb the exchange rate risk inherent in holding portfolios of assets denominated in different currencies when nominal exchange rates are volatile. These investors are referred to as *international arbitrageurs*.

In the face of shocks to the desired allocation of portfolios across assets denominated in different currencies, these international arbitrageurs are called on to absorb the nominal exchange rate risk inherent in such portfolio flows. That is, if households wish to reduce their holdings of euro bonds and increase their holdings of US dollar bonds, then, absent government intervention in bond supplies, international arbitrageurs must absorb that flow by increasing their holdings of euro bonds and decreasing their holdings of dollar bonds. In equilibrium, nominal exchange rates must move in a way that offers these international arbitrageurs a financial reward for taking on additional exchange rate risk in their portfolios. Here, this entails an immediate decline in the exchange rate value of the euro relative to the dollar so as to allow arbitrageurs to earn a high return when the euro returns over the long term to its prior level relative to the dollar. The research challenge is to integrate such a model of nominal exchange rate determination in asset markets with the behavior of international real relative prices in the markets for traded goods and the associated impact of these nominal exchange rate movements on the macroeconomy.

What Oleg and Dmitry achieve in [10] is such an integration. Specifically, they address the question of why, in the face of such financial market shocks to the nominal exchange rate, we do not simply see either volatile flows of traded goods across international boundaries or a relatively rapid response of inflation differentials across countries to restore purchasing power parity. They join the micro and the macro, bringing the insights from Oleg's prior work to bear. The exchange rate movements induced by financial shocks in the model do not result in significant macroeconomic responses in terms of reallocation of expenditure and output across countries, because of all the forces which Oleg previously studied that dampen the response of the terms of trade to exchange rate shocks. In this way, they offer a resolution of the exchange rate disconnect puzzle between the exchange rate and countries' underlying macroeconomies for the large major economies of the world.

In [11], Oleg and Dmitry take their model a step further to address the Mussa puzzle: they seek the causal mechanism through which a country might stabilize its international real relative prices by adopting a fixed exchange rate. In their model, international capital markets are particularly bad at dealing with nominal exchange rate risk because this risk is concentrated in a limited number of international arbitrageurs. When the perceived level of nominal exchange rate risk in the future is high, these arbitrageurs require large swings in their expected compensation for taking on more or less of this risk. These swings in the expected excess returns for international arbitrageurs correspond to large movements in the current level of the nominal exchange rate. In contrast, if a government can reduce the perceived level of nominal exchange rate risk in the future by adopting a fixed nominal exchange rate, these international arbitrageurs are happy to absorb large shocks to desired portfolios today with little or no compensation for risk because there is little such risk for them to be concerned about. In this case, these financial shocks affect neither the current exchange rate nor the macroeconomy because they are fully absorbed by private actors in international financial markets.

Papers [10] and [11] are quite recent, but they may have the greatest impact of Oleg's papers to date. Many countries in the world, particularly emerging market economies, have begun to experiment with unconventional policies to address shocks to capital flows together with exchange rate and macroeconomic volatility. Such policies go beyond the typical setting of nominal interest rates to include direct intervention in foreign exchange markets, measures to control cross-border flows of private capital, and macroprudential measures aimed at enhancing the stability of domestic financial sectors. As noted by Adrien and Gopinath (2020), policymakers worldwide are assessing these policy choices in a somewhat eclectic manner that does not rely on a clear analytical framework to assess how these policy tools should be used in an integrated way. Basu and others (2020a, b) develop a welfare theoretic framework to assess the optimal choice of multiple instruments in an integrated manner. Oleg's papers [10] and [11] are important contributions to this agenda.

The Impact of International Trade on Inequality

Early in his research, in [12], [13], and [14], Oleg studies the link between increased volume of international trade and inequality among workers. This question has been of intense policy interest, especially given the long post–World War II effort to reduce barriers to international trade through multilateral and bilateral agreements. Standard trade models typically address this question by looking at the impact of reduced barriers to international trade on workers with different observable characteristics—such as employment sector, education, experience, and occupation. In the data for the United States, however, much of the observed increase in inequality is "residual," in the sense that it is not accounted for by workers' observable characteristics. Oleg's work in this area is focused on understanding how trade affects both unemployment and this residual income inequality for workers.

In [14], coauthored with Elhanan Helpman, Oleg develops a model of the relationship between international trade and unemployment, a question not usually studied in standard models, which typically do not include the labor market frictions leading to unemployment.[3] This model serves as a framework for assessing the interaction of labor market frictions and impediments to trade in shaping welfare, trade flows, unemployment, and productivity.

In [13], coauthored with Helpman and Stephen Redding, Oleg extends this model to consider the interaction of trade and within-group inequality for employed workers. This model extends the standard Diamond-Mortensen-Pissarides model of search frictions, which results in a situation in which workers will be heterogenous in their unobserved ability. The next step is to embed this framework in a Melitz (2003) model of the selection of larger firms into international trade, because larger firms can afford to pay the fixed costs of participation in international markets. Taken together, this model explains that larger firms will screen applicants more intensively to employ higher-quality workers, pay these workers more, and select into participation in international trade. In the model, as in the data, firms that participate in international trade are larger, pay higher wages, and have higher labor productivity. Oleg and his coauthors develop novel results in this framework regarding the nonmonotonic relationship between increasing trade and inequality as barriers to international trade fall from a prohibitively high level to zero.

In [12], coauthored with Helpman, Redding, and Marc Muendler, Oleg takes this model to linked firm-worker data in Brazil. The authors first demonstrate that much of the inequality between workers in Brazil is driven by differences in wages across firms, consistent with their model. They then go on to estimate a structural model. This paper presented the first serious quantitative exercise evaluating the impact of trade cost reductions on within-group wage inequality.

[3]Davidson, Martin, and Matusz (1999) is an early study of the interaction of trade liberalization and unemployment generated by search frictions. Artuç, Chaudhuri, and McLaren (2010) study the dynamic impact of trade liberalization of wages when workers has switching costs between sectors.

Conclusion

The field of international macroeconomics has entered a new era of intellectual excitement based on advances both in theories and in the data and empirical strategies we have available to evaluate those theories. Oleg's research with his coauthors has fundamentally altered our understanding of the relationship between nominal exchange rates, prices of internationally trade goods, and macroeconomic fundamentals. His work has helped demystify several long-standing puzzles in the literature. In turn, this has led to a deeper understanding of monetary and exchange rate policy in open economies.

In addition to being a prolific scholar, Oleg is an exceptional teacher and collaborator. He has an infectious enthusiasm for everything he works on which comes through in the classroom and which makes him a great coauthor. Even as a student, Oleg handled the seminar jousting incredibly well and was never rattled. This quality to absorb constructive feedback while ignoring petty comments helped him professionally as an economist. His kindness and generosity has also made him a great mentor to students. He brings a boundless optimism and enthusiasm for economics to everything he does.

References

Adrien, Tobias, and Gita Gopinath. 2020. "Toward an Integrated Policy Framework for Open Economies." IMFBlog, July 13. https://www.imf.org/en/Blogs/Articles/2020/07/13/toward-an-integrated-policy-framework-for-open-economies.

Artuç, Erhan, Shubham Chaudhuri, and John McLaren. 2010. "Trade Shocks and Labor Adjustment: A Structural Empirical Approach." *American Economic Review* 100 (3): 1008–45.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing-to-Market, Trade Costs, and International Relative Prices." *American Economic Review* 98 (5): 1998–2031.

Backus, David K. and Gregor Smith. 1993. "Consumption and Real Exchange Rates in Dynamic Economies with Non-traded Goods." *Journal of International Economics* 35 (3–4): 297–316.

Basu, Suman S., Emine Boz, Gita Gopinath, Francisco Roch, and Filiz D. Unsal. 2020a. "A Conceptual Model for the Integrated Policy Framework." IMF Working Paper 2020/121.

Basu, Suman S., Emine Boz, Gita Gopinath, Francisco Roch, and Filiz D. Unsal. 2020b. "Integrated Monetary and Financial Policies for Small Open Economies." https://www.aeaweb.org/conference/2021/preliminary/paper/e844yY47.

Cavallo, Alberto, Brent Neiman, and Roberto Rigobon. 2014. "Currency Unions, Product Introductions, and the Real Exchange Rate." *Quarterly Journal of Economics* 129 (2): 529–95.

Davidson, Carl, Lawrence Martin, and Steven Matusz. 1999. "Trade and Search Generated Unemployment." *Journal of International Economics* 48 (2): 271–99.

De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, Nina Pavcnik. 2016. "Prices, Markups, and Trade Reforms." *Econometrica* 84 (2): 445–510.

Dixit, Avinash, and Victor Norman. 1980. *Theory of International Trade: A Dual, General Equilibrium Approach*. London: Cambridge University Press.

Engel, Charles. 1999. "Accounting for U.S. Real Exchange Rate Changes." *Journal of Political Economy* 107 (3): 507–38.

Engel, Charles. 2000. "Comment: The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" In *NBER Macroeconomics Annual 2000*, Vol. 15, edited by Ben S. Bernanke and Kenneth Rogoff, 403–11. Cambridge, MA: MIT Press.

Engel, Charles, and John H. Rogers. 1996. "How Wide Is the Border?" *American Economic Review* 86 (5): 1112–25.

Fama, Eugene F. 1984. "Forward and Spot Exchange Rates." *Journal of Monetary Economics* 14 (3): 319–38.

Feldstein, Martin S., and Paul R. Krugman. 1990. "International Trade Effects of Value-Added Taxation." In *Taxation in the Global Economy*, edited by Assaf Razin and Joel Slemrod, 263–82. Chicago: University of Chicago Press.

Flood, Robert P., and Andrew K. Rose. 1995. "Fixing Exchange Rates: A Virtual Quest for Fundamentals." *Journal of Monetary Economics*, 36 (1): 3–37.

Gabaix, Xavier, and Matteo Maggiori. 2015. "International Liquidity and Exchange Rate Dynamics." *Quarterly Journal of Economics* 130 (3): 1369–420.

Goldberg, Pinelopi K., and Rebecca Hellerstein. 2008. "A Structural Approach to Explaining Incomplete Exchange-Rate Pass-Through and Pricing-to-Market." *American Economic Review* 98 (2): 423–9.

Goldberg, Pinelopi K., and Rebecca Hellerstein. 2013. "A Structural Approach to Identifying the Sources of Local Currency Price Stability." *Review of Economic Studies* 80 (1): 175–210.

Gopinath, Gita. 2015. "The International Price System." Paper presented at the Inflation Dynamics and Monetary Policy Jackson Hole Economic Policy Symposium, Jackson Hole, WY, August 27.

Gopinath, Gita, and Roberto Rigobon. 2008. "Sticky Borders." *Quarterly Journal of Economics* 123 (2): 531–75.

Gopinath, Gita, Pierre-Olivier Gourinchas, Chang-Tai Hsieh, and Nicholas Li. 2011. "International Prices, Costs, and Markup Differences." *American Economic Review* 101 (6): 2450–86.

Gopinath, Gita, Emine Boz, Camila Casas, Federico Díez, Pierre-Olivier Gourinchas, and Mikkel Plagborg-Møller. 2020. "Dominant Currency Paradigm." *American Economic Review* 110 (3): 677–719.

Grossman, Gene M. 1980. "Border Tax Adjustments: Do They Distort Trade?" *Journal of International Economics* 10 (1): 117–28.

Jeanne, Olivier. 2000. "Comment: The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" In *NBER Macroeconomics Annual 2000*, Vol. 15, edited by Ben S. Bernanke and Kenneth Rogoff, 390–403. Cambridge, MA: MIT Press.

Jeanne, Olivier, and Andrew Rose. 2002. "Noise Trading and Exchange Rate Regimes." *Quarterly Journal of Economics* 117 (2): 537–69.

Kouri, Pentti J. K. 1976. "The Exchange Rate and the Balance of Payments in the Short Run and in the Long Run: A Monetary Approach." *Scandinavian Journal of Economics* 78 (2): 280–304.

Lerner, Abba P. 1936. "The Symmetry between Import and Export Taxes." *Economica* 3 (11): 306–13.

Lustig, Hanno, and Adrien Verdelhan. 2019. "Does Incomplete Spanning in International Financial Markets Help to Explain Exchange Rates?" *American Economic Review* 109 (6): 2208–44.

Meese, Richard A., and Kenneth Rogoff. 1983. "Empirical Exchange Rate Models of the Seventies: Do They Fit Out of Sample?" *Journal of International Economics* 14 (1–2): 3–24.

Melitz, Mark J., 2003. "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–725.

Mundell, Robert A. 1961. "A Theory of Optimum Currency Areas." *American Economic Review* 51 (4): 657–665.

Mundell, Robert A. 1963. "Capital Mobility and Stabilization Policy under Fixed and Flexible Exchange Rates." *Canadian Journal of Economics* 29 (4): 475–85.

Mussa, Michael. 1986. "Nominal Exchange Rate Regimes and the Behavior of Real Exchange Rates: Evidence and Implications." *Carnegie Rochester Conference Series on Public Policy 25*: 117–214.

Obstfeld, Maurice, and Kenneth Rogoff. 1996. *Foundations of International Macroeconomics*. Cambridge, MA: MIT Press.

Obstfeld, Maurice, and Kenneth Rogoff. 2000. "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?" In *NBER Macroeconomics Annual 2000*, Vol. 15, edited by Ben S. Bernanke and Kenneth Rogoff, 339–90. Cambridge, MA: MIT Press.

Rogoff, Kenneth. 1996. "The Purchasing Power Parity Puzzle." *Journal of Economic Literature* 34 (2): 647–68.

Rose, Andrew. 2000. "A Review of Some of the Economic Contributions of Robert A. Mundell, Winner of the 1999 Nobel Memorial Prize in Economics." *Scandinavian Journal of Economics* 102 (2): 211–22.

Journal of Economic Perspectives—Volume 37, Number 1—Winter 2023—Pages 241–248

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by e-mail at taylort@macalester.edu, or c/o Journal of Economic Perspectives, *Macalester College, 1600 Grand Ave., Saint Paul, MN 55105.*

Potpourri

Siwan Anderson delivered the Innis Lecture at the Canadian Economics Association last summer on "Unbundling Female Autonomy" (*Canadian Journal of Economics*, November 2022, 1671–701, https://onlinelibrary.wiley.com/doi/10.1111/caje.12628). Anderson writes (citations omitted): "Female empowerment is a multi-faceted concept that targets: improved female decision-making power in the household, reduction of violence against women, increased market and political opportunities, equal legal rights and dismantling gender-biased customs and norms. . . . Perhaps the classic argument in this area is that empowered women invest more in children. . . . In particular women want to, ceteris paribus, allocate relatively more of household resources to children's education and health than will

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.1.241.

men. Because both of these are crucial determinants of human capital formation and human capital formation is at least a proximate cause of economic development, development will be enhanced by factors that improve female autonomy (or a woman's outside option) relative to their husbands through the channel of increasing their control over the allocation of household resources. . . . One may conjecture that short-term policy interventions would be unlikely to significantly shift strongly embedded societal norms, given that many have persisted for centuries. However, emerging evidence suggests the contrary. For example, reserving seats for female politicians in rural areas of India has helped curtail negative stereotypes about women as local leaders. Television programs have been able to alter fertility preferences in multiple settings. Bursztyn et al. were able to adjust pre-determined individual Saudi male beliefs regarding the appropriateness of their wives' labour supply decisions by providing information on actual average male beliefs in their local geographical area. Regular secondary school class discussions, held amongst both boys and girls in India, were able to reshape some female negative attitudes and behaviours. . . . There is no reason, then, to expect that cultural changes in the currently developing world will mimic the paths followed in the West. . . . First, the timing of structural changes is different. Developing countries today experienced expansion of education and growth of the service sector at much lower levels of GDP per capita than when they took off in the West. Their legal contexts are also markedly different. Today's developing countries typically inherited the formal legal structures of their former colonists, which tend to be more progressive and favourable to women than the corresponding legal structures that prevailed at comparable levels of development in the West. At the same time, these formal legal structures often coexist in today's developing countries alongside extremely male-biased forms of customary law. Finally, there does not seem to be a massive shock to married women's labour supply, comparable to that occasioned by World War II, that could serve as a jolt to gender norms."

Jan Eeckhout reviews the evidence on "Dominant firms in the digital age" (UBS Center Public Paper #12, November 2022, https://www.ubscenter.uzh.ch/en/publications/public_papers/dominant-firms-in-the-digital-age.html). "The rise of dominant firms that we have seen during the advent of the digital age is built on cost-reducing and efficiency-enhancing innovations that create increasing returns to scale. This implies a winner-takes-all market with a dominant firm achieving a long-lasting monopoly position. And while monopoly is often associated with higher prices, most of these firms achieve this position by doing the opposite, that is lowering prices. They can do this because their innovations and investments lead to an even larger reduction in costs. And that is why the digital technology is so attractive for customers: technological innovation is the hero. But because costs decline more than prices due to scale economies, technological change is also the villain."

Brian R. Cheffins discusses "Getting Antitrust and History in Tune" (*Accounting, Economics, and Law: A Convivium*, published online March, 2, 2022, https://www.degruyter.com/document/doi/10.1515/ael-2021-0084/html). From the abstract: "Antitrust is high on the reform agenda at present, associated with calls to 'break up

big tech.' Proponents of reform have invoked history with regularity in making their case. They say reform is essential to reverse the baleful influence of the Chicago School of antitrust, which, in their telling, disastrously and abruptly ended in the 1980s a 'golden' era of beneficially lively antitrust enforcement. In fact, antitrust enforcement was, at best, uneven, from the early 20th century through to the end of the 1970s. As for the antitrust 'counter-revolution' of the late 20th century, this was fostered as much by fears of foreign competition and skepticism of government regulation as Chicago School theorizing. The pattern helped to ensure that the counter-revolution was largely sustained through the opening decades of the 21st century. This article, in addition to getting antitrust and history in tune by drawing attention to the foregoing points, provides insights regarding antitrust's future direction."

Gita Gopinath delivered the 2022 Martin S. Feldstein Lecture at the National Bureau of Economic Research on "Managing a Turn in the Global Financial Cycle" (*NBER Reporter*, October 2022, https://www.nber.org/reporter/2022number3/managing-turn-global-financial-cycle). "A key policy question therefore is how emerging and developing economies should respond to this tightening cycle that is driven to an important degree by rising US monetary policy rates. The textbook answer would be to let the exchange rate be the shock absorber. An increase in foreign interest rates lowers domestic consumption. By letting the exchange rate depreciate, and therefore raising the relative price of imports to domestic goods, a country can shift consumption toward domestic goods, raise exports in some cases, and help preserve employment. However, many emerging and developing economies find this solution of relying exclusively on exchange rate flexibility unsatisfying. This is because rising foreign interest rates come along with other troubles. They can trigger so-called 'taper tantrums' and sudden stops in capital flows to their economies. In addition, the expansionary effects of exchange rate depreciations on exports in the short run are modest, consistent with their exports being invoiced in relatively stable dollar prices. . . . Consequently, several emerging and developing economies have in practice used a combination of conventional and unconventional policy instruments to deal with turns in the global financial cycle. Unlike the textbook prescription, they not only adjust monetary policy rates but also rely on foreign exchange intervention (FXI) to limit exchange rate fluctuations, capital controls to regulate cross-border capital flows, and domestic macroprudential policies to regulate domestic financial flows. . . . Accordingly, to enhance IMF advice, David Lipton, the former first deputy managing director of the fund, championed the need to develop an Integrated Policy Framework that jointly examines the optimal use of conventional and unconventional instruments."

Every three years, the Bank of International Settlements conducts a survey of global over-the-counter foreign exchange markets. The *BIS Quarterly Review* includes five articles discussing results from the latest survey. For example, here's the abstract from "The global foreign exchange market in a volatile time," by Mathias Drehmann and Vladyslav Sushko (December 2022, https://www.bis.org/publ/qtrpdf/r_qt2212f.htm). "Turnover in global foreign exchange (FX) averaged more than

$7.5 trillion per day in April 2022 amid a volatile market environment. Compared with the previous BIS Triennial survey in 2019, trading volumes were higher because of greater activity in short-maturity FX derivatives and more inter-dealer trading. By contrast, trading with customers stagnated, mirroring a slowdown in international investment in 2022. A greater share of trading was executed via various bilateral methods, rather than via multilateral platforms that make prices available to all participants, implying that the transparency of the FX market may have decreased further."

The Congressional Budget Office publishes regular reports on inequality in the US economy: most recently, *The Distribution of Household Income, 2019* (November 2022, https://www.cbo.gov/publication/58781) and *US Household Wealth: 1989–2019* (September 2022, https://www.cbo.gov/publication/57598). From the income distribution report: "CBO's analysis compares Gini coefficients based on four different income measures: market income, income before transfers and taxes, income after transfers but before taxes, and income after transfers and taxes. . . . Between 1979 and 2019, income inequality as measured by the Gini coefficient for all four income measures rose. Increases in market income at the top of the distribution drove much of the rise in income inequality over that time. Of the four measures of income presented here, income inequality as measured by market income is the highest. Social insurance benefits, particularly Social Security and Medicare benefits, reduced income inequality relative to market income inequality. (Those benefits are included in income before transfers and taxes.) The progressive structures of means-tested transfers and federal taxes also reduced income inequality, but by smaller amounts than social insurance benefits did." From the wealth distribution report: "The total real wealth (that is, wealth adjusted to remove the effects of inflation) held by families in the United States tripled from 1989 to 2019—from $38 trillion in 2019 dollars (roughly four times the nation's gross domestic product, or GDP) to $115 trillion (about five times GDP). . . . The growth of real wealth over the past three decades was not uniform: Family wealth increased more in the top half of the distribution than in the bottom half. Families in the top 10 percent and in the top 1 percent of the distribution, in particular, saw their share of total wealth rise over the period. In 2019, families in the top 10 percent of the distribution held 72 percent of total wealth, and families in the top 1 percent of the distribution held more than one-third; families in the bottom half of the distribution held only 2 percent of total wealth."

Regulatory Economics

Robert S. Adler served as a Commissioner at the Consumer Product Safety Commission from 2009 to 2021, including as acting chair the last two years. He offers "Reflections of an Unapologetic Safety Regulator" (*Regulatory Review*, October 17, 2022, https://www.theregreview.org/2022/10/17/adler-reflections-of-an-unapologetic-safety-regulator/). He discusses what he calls the Great Safety Paradox:

"Paradoxically, the more successful regulators are in protecting the public, the less anyone notices. This paradox occurs because well-crafted safety rules do not raise prices or interfere with products' utility. In such cases, no one notices the improvement in safety. Most parents do not realize that the cribs they place their infants in no longer permit them to slip between the slats and strangle. Nor do they understand how much safer and less lead-laden their children's toys are. Similarly, most consumers will never recognize that their children no longer face being crushed by a garage door that unexpectedly closes on them or that infants do not suffocate in refrigerators because the doors can now be easily opened from within. Numerous government safety rules operate in a similar fashion, with life-saving benefits but little public recognition. . . . When health and safety agencies write a safety rule, they do so to eliminate or reduce deaths and injuries that consumers suffer in product-related accidents. The CPSC estimates that roughly 31,000 people die and 34 million people suffer product-related injuries every year. These deaths and injuries impose significant costs on the economy—roughly one trillion dollars annually. They do so first as medical costs and lost wages, then as higher premiums for health insurance—or higher taxes to pay for the uninsured. Moreover, product-related tragedies almost always result in a loss of economic productivity of the victims, not to mention the pain and suffering they experience. Accordingly, the argument that regulations necessarily impose new costs on society is not persuasive. The costs in the form of deaths and injuries are already there, and often they impose as much of a drag on the economy as any safety rule. . . . As former CPSC Commissioner R. David Pittle once said, 'it is far easier to redesign products than it is to redesign consumers.'"

Public Choice has published a six-paper symposium, plus an introduction, on "George Stigler's theory of economic regulation" (October 2022, https://link.springer.com/journal/11127/volumes-and-issues/193-1). Sam Peltzman contributed "Stigler's Theory of Economic Regulation After Fifty Years" (https://link.springer.com/article/10.1007/s11127-022-00996-0; ungated working paper version at https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=2625&context=law_and_economics). "[T]he Captured Regulator of 1971 is overstated but highly provocative. But without the provocation would we be here commemorating a fiftieth anniversary? . . . The capture theory does seem to fit some prominent cases, such as Stigler's motivating examples of truck regulation and occupational licensure. . . . We also have 20/20 hindsight of the proliferation of 'social regulation' that was underway when Stigler (1971) appeared. Environmental regulation is perhaps the most prominent example. Others include worker safety, the security of their pensions and consumer product safety. By some measures this regulatory expansion was, and remains, historically unprecedented. Typically social regulation cut across many industries. And it was invariably resisted by those industries. On the other side, deregulation of industries like transportation and securities brokerage surfaced in the late 1970s amidst significant industry resistance. Then more recently we get 'reverse capture,' where the industry is created by the regulator—as in renewable energy, biofuels and the like. None of these developments

seem contemplated by the capture theory. . . . The distinction I want to pursue is between the creation . . . and the output (design and operation) of regulatory bodies. Even casual history suggests that these often respond to different political forces and interest groups. In particular, the industry often—perhaps mainly—resists the establishment of regulation. The affected industries resisted the consumer reforms of the Progressive Era, the labor reforms of the New Deal and the social regulation of the 1970s. But, once confronted with the reality of the regulation, the industry interest usually plays a prominent role in what these agencies do."

The *Harvard Journal of Law & Public Policy* has collected five papers for "A Symposium on Regulatory Budgeting" (Summer 2022 online-only Per Curiam issue, https://www.harvard-jlpp.com/a-symposium-on-regulatory-budgeting/). For example, Andrea Renda discusses the "Regulatory Budgeting: Inhibiting or Promoting Better Policies?" She writes: "Over the past two decades, several governments have introduced tools to incentivize regulators to become more aware of the costs they impose on businesses and citizens when they propose new rules. . . . [G]overnments of various political orientations have introduced forms of regulatory budgeting, which require administrations to identify, every time they introduce new regulation entailing significant regulatory costs, provisions to be repealed or revised, so that the net impact on overall regulatory costs is (at least) offset. These rules are generically referred to as 'One-In-X-Out' (OIXO). . . . In their most common form of 'One-In-One-Out' (OIOO), these rules amount to a commitment not to increase the estimated level of burdens over the chosen time-frame. The OECD refers to these commitments as 'regulatory offsetting.' . . . There are at least twenty countries in the world that have adopted an OIXO rule. These include ten EU member states (Austria, Finland, France, Germany, Hungary, Italy, Latvia, Lithuania, Spain and Sweden) as well as Canada, Mexico and Korea. In the past, three countries have had a similar rule in place (Denmark, the UK, and the United States), but later decided to gradually phase it out . . . . Four other countries were reportedly introducing similar regulatory budgeting systems in 2020: Poland, Romania, Slovakia, Slovenia. . . . If carefully designed, regulatory budgeting rules are not incompatible with an ambitious policy agenda. In Germany, for example, the OIOO rule was adopted in a context in which by ambitious programs such as Energiewende are in place, and a systematic scrutiny of the impact of new legislation on sustainable development is carried out. In France, the government uses the OI2O rule but at the same time adopts ambitious proposals in terms of social and environmental benefits. In short, there is no incompatibility per se between the adoption of a cost reduction or regulatory budgeting system and an ambitious regulatory and policy agenda in the social and environmental domain."

The US Environmental Protection Agency has published its estimate of a social cost of $190 per metric ton of carbon emissions ("Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances," September 2022, https://www.epa.gov/system/files/documents/2022-11/epa_scghg_report_draft_0.pdf). Kevin Rennert and Brian C. Prest offer a blessedly readable overview in "The US Environmental Protection Agency Introduces a New Social Cost of Carbon

for Public Comment" (Resources for the Future, November 15, 2022, https://www.resources.org/common-resources/the-us-environmental-protection-agency-introduces-a-new-social-cost-of-carbon-for-public-comment/). "In its sensitivity analysis, EPA updates each of the four major steps of SCC [social cost of carbon] estimation: socioeconomic projections, climate modeling, translation to economic damages, and economic discounting. In doing so, the agency draws heavily on peer-reviewed and published work from the SCC Initiative, a multi-institution collaborative effort led by RFF and the University of California, Berkeley. This work includes the RFF-Berkeley Greenhouse Gas Impact Value Estimator (GIVE) model, which was recently published in the journal *Nature*." To give an idea of the sensitivities here: The central EPA estimate for the social cost of carbon—$190 per metric ton—uses a 26 percent discount rate. But the estimate would be $120/ton with all the same underlying estimates and a discount rate of 2.5 percent, and $340/ton with all the same estimates and a discount rate of 1.5%. Rennert and Prest also write: "In a major step forward for transparency, the computer code used for the sensitivity has been built using the open-source Mimi software platform (another output of the SCC Initiative), making the code free and easily accessible to download, replicate, and evaluate."

Discussion Starters

The environmental organization Greenpeace challenges the practicality and benefits of plastics recycling in "Circular Claims Fall Flat Again" (October 24, 2022, https://www.greenpeace.org/usa/reports/circular-claims-fall-flat-again/). "Mechanical and chemical recycling of plastic waste has largely failed and will always fail because plastic waste is: (1) extremely difficult to collect, (2) virtually impossible to sort for recycling, (3) environmentally harmful to reprocess, (4) often made of and contaminated by toxic materials, and (5) not economical to recycle. Paper, cardboard, metal, and glass do not have these problems, which is why they are recycled at much higher rates. Due to toxicity risks, post-consumer recycled plastic from household waste is not being produced at commercial scale for food-grade uses globally or in the U.S., and likely never will be."

Jennifer Randles discusses "Fixing a Leaky U.S. Social Safety Net: Diapers, Policy, and Low-Income Families" (*RSF: The Russell Sage Foundation Journal of the Social Sciences*, August 2022, 8:5, 166-183, https://www.rsfjournal.org/content/8/5/166). "Diaper need—lacking enough diapers to keep an infant dry, comfortable, and healthy—affects one in three mothers in the United States, where almost half of infants and toddlers live in low-income families. Diaper need . . . exacerbates food insecurity, can cause parents to miss work or school, and is predictive of maternal depression and anxiety. When associated with infrequent diaper changes, it can lead to diaper dermatitis (rash) and urinary tract and skin infections. Infants in the United States will typically use more than six thousand diapers, costing at least $1,500, before they are toilet trained. Cloth diapers are not a viable alternative

for most low-income parents given high start-up and cleaning costs and childcare requirements for disposables. Many low-income parents must therefore devise coping strategies, such as asking family or friends for diapers or diaper money; leaving children in used diapers for longer; and diapering children in clothes and towels. Low-income parents also turn to diaper banks, which collect donations and purchase bulk inventory for distribution to those in need and usually provide a supplemental supply of twenty to fifty diapers per child per month. In 2016, the nation's more than three hundred diaper banks distributed fifty-two million diapers to more than 277,000 children, meeting only 4 percent of the estimated need. Many of those who seek diaper assistance live in households with employed adults who have missed work because of diaper need."

Kathleen Fear, Carly Hochreiter, and Michael J. Hasselberg suggest "Busting Three Myths About the Impact of Telemedicine Parity" (*NEJM Catalyst*, October 2022, 3:10, https://catalyst.nejm.org/doi/full/10.1056/CAT.22.0086). "Three beliefs—that telemedicine will reduce access for the most vulnerable patients; that reimbursement parity will encourage overuse of telemedicine; and that telemedicine is an ineffective way to care for patients — have for years formed the backbone of opposition to the widespread adoption of telemedicine. However, during the Covid-19 pandemic, institutions quickly pivoted to telemedicine at scale. Given this rapid move, the University of Rochester Medical Center (URMC) had a natural opportunity to test the assumptions that have shaped prior discussions. Using data collected from this large academic medical center, UR Health Lab explored whether vulnerable patients were less likely to access care via telemedicine than other patients; whether providers increased virtual visit volumes at the expense of in-person visits; and whether the care provided via telemedicine was lower quality or had unintended negative costs or consequences for patients. The analysis showed that there is no support for these three common notions about telemedicine. At URMC, the most vulnerable patients had the highest uptake of telemedicine; not only did they complete a disproportionate share of telemedicine visits, but they also did so with lower no-show and cancellation rates. . . . Importantly, this access does not come at the expense of effectiveness. . . . As the pandemic continues to slow down, payers may start to resist long-term telemedicine coverage based on previous assumptions. However, the experience at URMC shows that telemedicine is a critical tool for closing care gaps for the most vulnerable patient populations without lowering the quality of care delivered or increasing short-term or long-term costs."

Why an AEA Ombudsperson?

- Informal
- Confidential
- Impartial
- Independent

If you have experienced or witnessed incidents of harassment or discrimination, the AEA Ombudsperson can help guide you.

The AEA Ombudsperson enables AEA members to speak anonymously with a neutral third party about potential violations of AEA policies and determine if a formal or an informal report is appropriate. Whether the member has been personally impacted by misconduct or was witness to it, the ombudsperson is a confidential resource who can advise the member about additional resources and possible next steps. No action is ever taken without the complainant's authorization.

Need Assistance? Please Reach Out.

Leto Copeley, AEA Ombudsperson

aeaombuds@whiteandstradley.com

919-844-0400

https://copeleylaw.com/aea-ombudsperson-contact/

White & Stradley, PLLC
3105 Charles B Root Wynd
Raleigh, NC 27612

Formal Complaint?

Contact with the AEA Ombudsperson does not constitute official notice to the AEA or trigger a formal complaint. Formal complaints are filed through the AEA and are reviewed and investigated by the AEA Ethics Committee. The AEA Ombudsperson can provide guidance about the formal process and may separately advise the Association about observations of emerging patterns, repeat perpetrators, or systemic indicators.

Find more information about these policies and the AEA Ombudsperson at

www.aeaweb.org/ombudsperson

AMERICAN ECONOMIC ASSOCIATION

The AEA is dedicated to improving the climate of the economics profession by addressing harassment and discrimination, which are in violation of AEA policies and the AEA Code of Conduct.

SUPPORTING DIVERSITY IN ECONOMICS

The Committee on the Status of Minority Groups in the Economics Profession (CSMGEP) was established by the American Economic Association (AEA) in 1968 to increase the representation of minorities in the economics profession, primarily by broadening opportunities for the training of underrepresented minorities.

CSMGEP Programs

- Summer Economics Fellows Program
- Mentoring Program
- Summer Training Program
- Initiatives for Diversity and Inclusion
- The Minority Report Newsletter

www.csmgep.org

AEA INITIATIVES FOR DIVERSITY AND INCLUSION

For more details and information regarding how to apply for AEA diversity initiatives, please visit **www.aeaweb.org/go/diversity-initiatives**

The American Economic Association is committed to the continued improvement of the professional climate in economics. In cooperation with key committees, the Association has launched several initiatives to support and promote diversity and inclusion in our profession.

1 AEA Award for Outstanding Achievement in Diversity and Inclusion

This annual award will recognize departments and organizations that demonstrate outstanding achievement in diversity and inclusion practices. Focus will be on those applicants that take productive steps to establish new programs and procedures to create an inclusive environment, and to increase the participation of underrepresented racial/ethnic minorities, women, and LGBTQ+ individuals.

2 Departmental Seed Grants for Innovation in Diversity and Inclusion

These grants, in amounts up to $5,000, will be awarded to economics departments to help establish new bridge programs or training programs for underrepresented minorities (URM). For example, a department might create a mentoring program for URM graduate or undergraduate students, create opportunities for URM students to do meaningful research assistant work, or start a program allowing URM students who need additional preparation for graduate school to take a lighter class load in the first year or to take core economics courses over two years.

3 The Andrew Brimmer Undergraduate Essay Prize

Thanks to the generosity of an anonymous donor, this paper prize has been established in honor of Andrew Brimmer, the first African American to serve on the Board of Governors of the Federal Reserve. The annual award will be presented to an undergraduate student at a US-based institution of higher learning majoring in economics, political science, public policy, or related fields for the best essay on the "economic well-being of Black Americans." The winner will receive a check for $1,000 and a plaque from the president of the AEA.

4 URM Travel Grants

This award is open to junior economics faculty members from traditionally underrepresented groups in the economics profession. The grants will advance career and professional development by defraying the costs of travel, lodging, and conference registration to attend the annual ASSA Meeting.

5 Small Group Breakfast Meeting for URM

Each year at the ASSA Meeting there will be a breakfast held with scholars from underrepresented minorities and prominent economists in attendance. The goal is to allow URM scholars access to AEA journal editors, executive board members, thought leaders in specific areas of economics, or other economists for the purpose of addressing issues of access to journals, conferences, and networks that are often out of reach for URM scholars.

6 Professional Development Grant for URM

This $2,000 grant was established to help advance the career and professional development of URM in the field of Economics. The award is open to eligible junior economics faculty members. Entrants to the essay competition should detail their research and how it relates to economics education.

These initiatives are another important step in helping make our field accessible and welcoming to anyone with the interest and ability to make a career in it. Please help us share this information throughout the profession so we can all work together and continue to improve.

SUPPORTING LGBTQ+ INDIVIDUALS IN THE ECONOMICS PROFESSION

The Committee on the Status of LGBTQ+ Individuals in the Economics Profession (CSQIEP) was established by the American Economic Association in 2019 (following three years of support of a related Ad Hoc Working Group on LGBTQ+ Economics) to address issues facing LGBTQ+ economists.

The Committee supports the AEA by offering recommendations on Best Practices concerning sexual orientation and gender identity issues in economics, supports LGBTQ+ economists through a variety of professional development and mentoring opportunities, and works in tandem with CSMGEP, CSWEP, and related AEA committees to advance equity, inclusion, and diversity throughout the economics profession.

CSQIEP Activities

- Mentoring Program
- Virtual Seminar Series
- Pink Papers at the AEA Annual Meeting
- Outstanding Paper Award
- Maynard's Notes LGBTQ+ Newsletter

@LGBTQ_Econ

www.csqiep.org

AEA Training Opportunity!

EDUCATE Workshop

EXPANDING DIVERSITY in UNDERGRADUATE CLASSES with ADVANCEMENTS IN (THE) TEACHING (OF) ECONOMICS

Overview

This workshop provides opportunities for instructors of undergraduate courses to take part in course design activities and experience pedagogical strategies that will engage all of the students they teach. Attendees will have opportunities to identify learning objectives that focus on the students' ability to "do economics" and to participate in pedagogical practices that enable students to be active participants inwonomic analysis.

All accepted applicants are expected to fully engage with each of the three phases of the overall program including constructing learning objectives, studying pedagogical practices that are collaborative and inclusive including cooperative learning, engaging lectures, data integration, and classroom experiments, and integrating lessons learned into their own courses. Participants will be provided opportunities to share their work at the 2024 CTREE meeting.

Participants will engage with issues of diversity and inclusion throughout the workshop including opportunities to think critically about course goals and learning outcomes, their relationship to pedagogical choices and assessment, and how such decisions might have disparate effects on those of different races, genders, and ethnicities. In addition, attendees will learn how to discuss the sensitive topics that are an important part of the economics classroom.

Eligibility

Applicants must be scheduled to teach during the following fall and spring semesters. Preference will be given to those less than 6 years since PhD. To meet the goals of increasing diversity in the profession, the cohort will be chosen to represent a diverse set of institutions and instructors.

Cost

Accepted applicants must make an electronic payment of $100 to confirm their spot in the workshop. Housing (2 nights) and meals during the workshop will be covered by the AEA under the condition that the accepted applicant attends ALL workshop related activities.

The application portal opens in January 2023 with a rolling acceptance procedure that continues until all workshop slots are filled. Workshop details and the application portal are available at https://www.aeaweb.org/go/educate-workshop.

EDUCATE Workshop

(A face-to-face and Canvas-supported course)

June 2-3-4, 2023
Portland, OR

For more information go to
https://www.aeaweb.org/go/educate-workshop

EDUCATE is sponsored by the AEA Outreach Task Force and the AEA Committee on Economic Education.

THE COMMITTEE ON THE STATUS OF WOMEN IN THE ECONOMICS PROFESSION

CSWEP

Advancing the Status of Women in the Economics Profession

- Publishes an annual survey of the representation of women in economics
- Offers mentoring workshops for junior faculty, graduate students, and early-career researchers outside of academia
- Conducts programs at the AEA and at meetings of other professional associations
- Co-sponsors the summer economics fellows program
- Hosts career development webinars
- Celebrates outstanding economists who have furthered the status of women in the profession
- Publishes *CSWEP News*. Free newsletter subscription available at **info@cswep.org**

www.CSWEP.org **@AEACSWEP**

JOE NETWORK

The career network designed by economists—for economists.

The AEA's JOE Network is the preferred hiring tool for the economics job market, matching qualified candidates with relevant economics positions in leading institutions. Over 1,700 positions are filled each year from a pool of 5,100+ candidates.

Job Seekers

Showcase your profile to hundreds of hiring managers and quickly apply to your preferred positions. The JOE Network allows you to securely request reference letters and share job market materials.

Employers

Target search requirements to locate your preferred candidates, set up hiring committee access, and easily manage job postings and applications.

Faculty Letter-Writers

Reply to reference requests with default or customized letters, assign proxies, and easily monitor task completion. Over 150,000 requests are fulfilled on the JOE Network each year.

Use the JOE Network as your comprehensive tool for the economics job market!

AMERICAN ECONOMIC ASSOCIATION

aeaweb.org/JOE

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary access to JEP articles, go to the AEA website: http://www.aeaweb.org. Change of address notice must be received at least six weeks prior to the publication month.

The Journal of

Economic Perspectives

Winter 2023, Volume 37, Number 1

Symposia

Trade Sanctions and International Relations

T. Clifton Morgan, Constantinos Syropoulos, and Yoto V. Yotov, "Economic Sanctions: Evolution, Consequences, and Challenges"

Marco Cipriani, Linda S. Goldberg, and Gabriele La Spada, "Financial Sanctions, SWIFT, and the Architecture of the International Payment System"

Monetary Policy

Anil K Kashyap and Jeremy C. Stein, "Monetary Policy When the Central Bank Shapes Financial-Market Sentiment"

Michael D. Bauer, Ben S. Bernanke, and Eric Milstein, "Risk Appetite and the Risk-Taking Channel of Monetary Policy"

Alan S. Blinder, "Landings, Soft and Hard: The Federal Reserve, 1965–2022"

Alisdair McKay and Christian K. Wolf, "Monetary Policy and Inequality"

Hispanic Americans

José Fernandez, Mónica García-Pérez, and Sandra Orozco-Aleman, "Unraveling the Hispanic Health Paradox"

Francisca M. Antman, Brian Duncan, and Stephen J. Trejo, "Hispanic Americans in the Labor Market: Patterns over Time and across Generations"

Gordon Hanson, Pia Orrenius, and Madeline Zavodny, "US Immigration from Latin America in Historical Perspective"

Article

Andrew Atkeson and Gita Gopinath, "Oleg Itskhoki: 2022 John Bates Clark Medalist"

Feature

Timothy Taylor, "Recommendations for Further Reading"

AMERICAN ECONOMIC ASSOCIATION