The Journal of

Economic Perspectives

A journal of the
American Economic Association

Summer 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), Summer 2023, Vol. 37, No. 3. 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 3 • Summer 2023

Statement of Purpose

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

Policy on Data Availability

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

Policy on Disclosure

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

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

How Far Goods Travel: Global Transport and Supply Chains from 1965–2020

Sharat Ganapati and Woan Foong Wong

The integration of countries and industries into global supply chains depends on cheap and efficient transport. We show the evolution of transport use and costs over the last 55 years and establish their implications for international trade and global supply chains. To set the stage, consider a concrete example: the change in the manufacture of telephones from a century ago to the present day.

Built in 1905, the Western Electric Hawthorne Works factory in the Chicago suburb of Cicero, Illinois, manufactured 43,000 varieties of telephone apparatus for the parent Bell telephone monopoly (Weber 2002; Schlagheck and Lantz 2014). It employed 40,000 people who worked in over 100 buildings. Even with a transcontinental railroad system, transport costs were substantial, and excessive back-and-forth transport links were not common. As such, while this factory did source a few raw materials like Bakelite, rubber, and metal from remote locations, it manufactured many intermediate components internally—such as vacuum tubes in the early days and transistors later—before distributing finished telephone equipment across the country. This manufacturing complex effectively made telephone handsets in a single location for the entire United States.

The factory operated until 1986, and large portions of the grounds were dynamited in 1994 to build a suburban shopping complex (Pelton 1994). In the age of

■ *Sharat Ganapati is Assistant Professor of International Economics, Georgetown University, Washington, DC, and Faculty Research Fellow, National Bureau of Economic Research, Cambridge, Massachusetts. Woan Foong Wong is Assistant Professor of Economics, University of Oregon, Eugene, Oregon. Their email addresses are sg1390@georgetown.edu and wfwong@uoregon.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.3.3.

globalization and low transport costs, a vertically integrated factory in a high-cost location no longer made financial sense.

The supply chain for the modern smartphone is quite different. The research and design activities for Apple's iPhone take place in the United States, with further engineering in the United States and Taiwan (including within its largest partner, Foxconn). Production directly involves 43 countries in six continents in addition to any further upstream manufacturers; key components are manufactured in Japan, Korea, Taiwan, and China, with final assembly in China and India (Dedrick and Kraemer 2017; Petrova 2018). Apple's direct subcontractors do not manufacture many of the components used and only assemble the final product before shipment around the world. With the exact mix depending on the model, components such as memory, microprocessors, optics, batteries, and screens are manufactured in both nearby Asian countries like South Korea, Taiwan, Japan, Malaysia, and Vietnam, or even in the United States, Mexico, or European Union.

The supply chain for Apple's iPhone is not unusual. In Samsung's smartphone production process, design takes place in South Korea, manufacture of key components takes place in South Korea, Japan, and the United States, and the final assembly takes place in Korea, Vietnam, China, India, Brazil, and Indonesia (Dedrick and Kraemer 2017). These locations are connected by frequent and reliable shipping networks. The expansive use of global networks by companies like Apple and Samsung is a function of declining transportation costs (discussed in this journal by Hummels 2007). In 1890, it cost nearly $200 per ton (in 2020 dollars) to ship goods from California to Europe. A century later, the cost would be less than $2 per ton using a standard bulk ship (Harley 1988).

In this essay, we first set the stage with some facts and patterns. We document the dramatic rise in global transport use from 1965 to 2020, as measured either by weight in ton-kilometers traveled as is standard in the transportation literature, or by value in dollar-kilometers as is standard in the trade literature. After accounting for economic growth, real transport use per unit of final consumption has more than doubled over this period, increasing by 100 percent by weight and 160 percent by value. We also document that while real transport use by weight continued increasing after the 2007 Great Recession, real transport use by value has substantially declined. Second, we establish trends on global transport costs and show that they have declined over the last half century by 33–39 percent and 48–62 percent, by weight and value respectively. Third, we consider the factors that contributed to the transport use increase, especially the participation of emerging economies; in particular, since 1990, China has accounted for the entirety of the relative global transport use increase by weight. More generally, trade over longer distances, more than 5,000 kilometers, accounts for most of the transport use increase, compared to shorter distance trade. Transport use increases by weight are also driven by natural resources and raw materials, while downstream manufactured goods drive the increase by value. Fourth, we consider some key technology and infrastructure changes which contributed to these changes in transport use, including container and jet airplane technology, economies of scale in shipping, and innovations in logistics management like "just-in-time" deliveries.

Transport Use over Time

World trade has exploded since the end of World War II, accounting for an increasing share of production and consumption. The World Trade Organization (2023) reports that world trade is 43 times larger by volume in 2021 than in 1950.

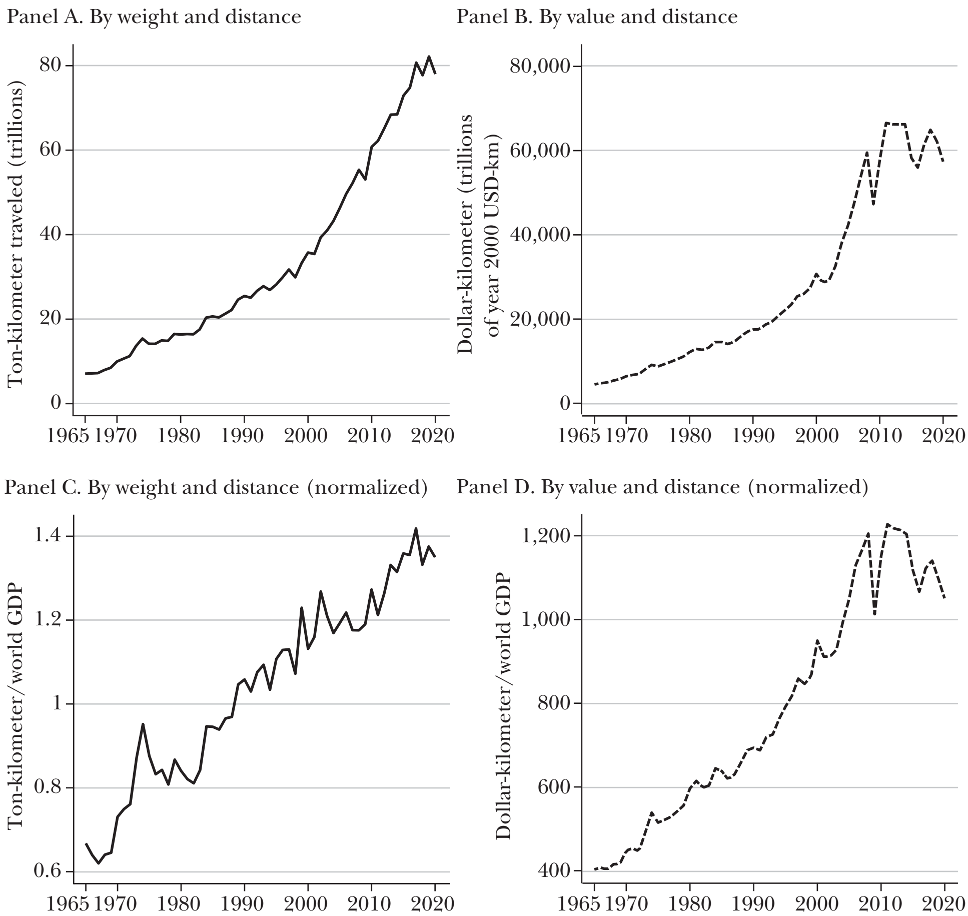
We examine this increase in global trade and its link to increases in the use of global transport services—not only to ship more goods, but to ship them further as well. Specifically, we can think of the use of transportation services as primarily consisting of two components: the amount of goods that are transported and how far these goods are transported (for an algebraic presentation of the approach described in this section, see online Appendix A). The first component, international trade flows, is captured using conventional trade statistics. The second component is important to incorporate, as goods that are shipped further require more transportation services. This transport use measure captures what is often missing in traditional trade measures—the role of distance. If trade increases, but only between nearby countries, then the transportation use increase will be mostly driven by the first component—trade flows. But if trade between distant locations increases, then both components contributing to transportation needs will increase. Including distance directly captures transport use.

We measure transportation usage in two ways. The first method uses the *weight* of transported goods, multiplied by the distance traveled. Most transportation costs are primarily priced in either weight or volume (Hummels and Skiba 2004; Irarrazabal, Moxnes, and Opromolla 2015; Wong 2022). Bulk cargo transport costs are typically measured in tons, while containers are priced by volume as measured in "twenty-foot equivalent" units or TEUs. This weight-based measure is more reflective of goods with lower value-per-weight ratios, such as grain, coal, ore, or petroleum products. The second method uses the *value* of goods multiplied by the distance traveled. This value measure places emphasis on the transport of goods with higher value-per-weight, like machinery, automobiles, and electronics. In both measures, multiplying by distance gives a better sense of transportation use, which is often missing in traditional trade measures that sum exports and imports.

In Figure 1, the top two panels demonstrate that international transportation usage has increased from 1965 to 2020. We use all trade between origin and destination for 200 countries, measured in tons and in dollars (converted to 2000 US dollars), based on the National Bureau of Economic Research-United Nations (NBER-UN) Comtrade and Centre d'Informations Internationales' Base pour l'Analyse du Commerce International (CEPII BACI) databases (Feenstra et al. 2005; Gaulier and Zignago 2010; Conte, Cotterlaz, and Mayer 2021). For distance between countries, we use the population-weighted, as-the-crow-flies distance as an approximation, because we do not observe the specific route or transportation of goods.[1]

[1] As-the-crow-files distance is also known as the "great circle" or haversine distance. Some caveats are worth noting. While data on the value of trade is available over our entire sample period, weight data is

Figure 1
Transport Use, 1965–2020



Sources: BACI, the World Input-Output Database (WIOD), UN-NBER-Comtrade, Penn World Table (PWT), and associated output deflators.
Notes: Panels A and B measure the distance shipped of goods, multiplied by metric tons and real year 2000 US dollars respectively (for further details see equation [1] in online Appendix A). While data on the value of trade is available over our entire sample period, weight data is only widely available after 2000. We impute weight data from prices using a variety of data sources. See online Appendix B for full details. Panels C and D normalize the top two figures relative to the sum of the gross domestic product of all countries to calculate real transport use (for further details see equation [2] in online Appendix A). All monetary values are converted to year 2000 US dollars.

only widely available after 2000. We impute the weight from prices using World Input-Output Database (WIOD) price index data and BACI price/weight data from 1995–2000. For natural resources from 1970–1985, data from the US Energy Information Administration (EIA) and the Organization of the Petroleum Exporting Countries (OPEC) are used to impute weights from prices. Full data details are in the online Appendix.

Figure 1, panel A, shows that transport usage by weight increased more than tenfold, from about 7.1 trillion ton-kilometers in 1965 to about 78 trillion ton-kilometers in 2020. Figure 1, panel B, shows a 14-fold increase when measuring transport use in value terms—from 4,000 trillion dollar-kilometers in 1965, to 67,000 trillion dollar-kilometers in 2011, and a modest decline to 57,000 trillion dollar-kilometers in 2020.

The two trends mirror each other from 1965 to the Great Recession. After 2008, the weight measure of transport usage (in ton-kilometers) continues its rapid growth, but trade use as measured in dollar-kilometers declines. Potential explanations include less trade of higher value-to-weight goods and shorter transport distances of these goods. We will revisit the diverging trends in transport usage by weight and value later.

We could employ more direct measures of transportation use, but such measures are often only available for subsets of countries, for shorter time periods, and for specific modes of transportation. As an example, available Census Bureau trade data that breaks down US imports and exports by air, ocean, and containers only starts in 1992 and covers only trade outside the North American Free Trade Agreement (NAFTA). Similarly detailed data is not available for most other countries. Our approach allows us to measure transportation use for more broadly (for 200 countries) and over a much longer period (55 years).

Since the world economy is growing rapidly during this time, we next account for this growth by normalizing total transportation usage by real global GDP. By normalizing relative to final output (as measured in GDP), this real transport use also captures the cumulative distance traveled by intermediate inputs in production, in addition to the distance traveled by the final good to its ultimate destination for consumption.

Returning to the telephone handset example, when calculating its real transport use in 1965, our method includes the distance traveled for raw materials from Asia, South America, and Central America to the United States for use in intermediate input production at the Western Electric Factory. Subsequent final assembly all happens at the same location and adds no further distance in terms of international trade (Western Electric 1938).

Today, with smartphones such as Apple's iPhone, raw materials from Brazil and Africa get shipped to Vietnam and the European Union to be made into plastic and silicon, which are then sent to Taiwan and South Korea to be manufactured into memory modules. These modules, along with a variety of other components, such as microprocessors and LCD screens, are then combined in India and China for products that are shipped globally for final consumption. Everything is shipped around the world, both low weight-to-value raw materials and high weight-to-value final assemblies (Dedrick and Kraemer 2017). Real transport use for smartphones will be much higher compared to the telephone handset due to its production process taking place in many more countries and to both the raw materials and final assemblies being transported much further distances. By normalizing transport use relative to real global GDP, we can show the transport use changes after accounting for the growing world economy over this period.

As shown in Figure 1, panel C, our measure of real transport use in weight has more than doubled over the past 50 years, from 0.67 ton-kilometer per dollar of real GDP in 1965 to 1.35 ton-kilometer per dollar of real GDP in 2020. As previously mentioned, the weight-based measure of trade is more reflective of the transport of raw materials. When using value measures in Figure 1, panel D, this increase in normalized transport use is even larger than the weight measure—tripling from 1965 to 2007, before declining nearly 20 percent from the peak. As mentioned previously, this value measure is a closer approximation for the transport of final goods. As an example, more telephones are traded globally today than yesterday. In both bottom panels, transport use is increasing above and beyond the growth in the world economy.

We now compare our measures of transportation usage to conventional trade statistics. In Figure 2, we plot our two normalized transportation usage measures in weight-distance and value-distance from the bottom panels of Figure 1 against the growth of more conventional trade measures, trade values, and trade weights as a share of global output.[2] The more conventional trade measures do not account for distance, and so this comparison allows us to highlight the role of distance over this period—are more goods being shipped to countries that are further apart?

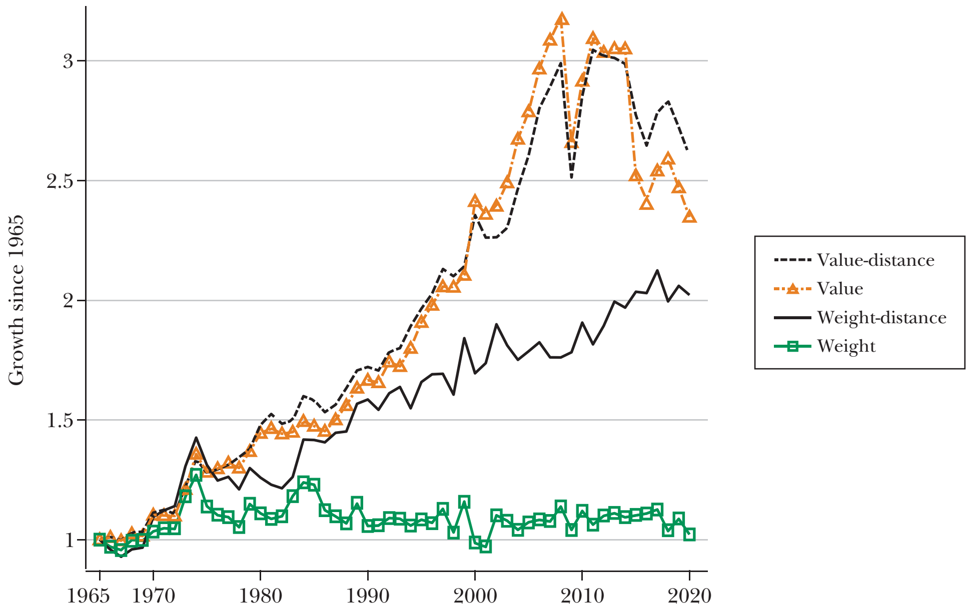
We emphasize three themes that emerge. First, when trade is measured by value, the growth of normalized transport usage in dollar-distance (gray dashed line) echoes the growth of trade value as share of global output (orange triangles)—more than tripling by 2007 relative to 1965, before decreasing after the Great Recession (Eaton et al. 2016). From 1965 to 2020, trade has increased by 2.5 times. In value, goods are being shipped more to both nearby and distant locations.

Second, when trade is measured by weight, the growth of normalized transport usage in weight-distance (black solid line) is quite different from with the growth of trade weight as a share of global output (green squares). These two series diverge early in our sample period. As mentioned before, transport usage in weight-distance steadily increases at a slower rate and more than doubles by 2020 (compared to the dollar-distance measure). However, as a share of the global economy, the aggregate amount of tonnage shipped has stayed relatively constant from 1965 to 2020 at around 0.24 to 0.26 shipped tons per $1,000 of real world GDP. Relative to the global economy, nations are not trading significantly more goods by weight. However, when nations do trade these goods, they are transported over increasingly further distances.

Third, the growth in our normalized trade statistics using ton-kilometer and dollar-kilometer parallel each other until 1990. After 1990, the growth in the normalized value measure accelerates through 2007 and then subsequently collapses. Meanwhile, growth in the normalized weight measure of trade continues to rise steadily throughout this period, largely unaffected by the 2008 recession.

[2] We can follow Johnson and Noguera (2012) to reframe the value portion as value added. This dampens growth by 10–25 percent but does not alter the qualitative features of the comparison.

Figure 2
Transport Use, 1965–2020



Sources: BACI, WIOD, UN-NBER-Comtrade, PWT, and associated output deflators.
Notes: All series are normalized with respect to its 1965 value and can be read as growth since 1965. The black line is the real or normalized transportation use measured in ton-kms from Figure 1, panel C. The dashed gray line is the real or normalized transportation use measured in dollar-kms from Figure 1 panel D. The green squares are the total weight shipped relative to World GDP. The orange triangles are the value shipped relative to World GDP. All monetary values are converted to year 2000 USD. In addition to the data from the aggregate measures of trade and transportation from Figure 1, WIOD data is used for country-level GDP from 1965–2014, and the Penn World Table for 2015–2019 (Feenstra, Inklaar, and Timmer 2015), and UN statistics Division (2022) for GDP data from 2015–2020. See the online Appendix for full details.

In the last decade or so, nations have cut back purchasing higher value-to-weight goods like electronics—including devices such as Apple iPhones—relative to overall consumption, but trade in lower value-to-weight goods, which are more reflective of raw materials like coal or oil, continue to grow.

These points highlight a central tension at the intersection of the two fields of transportation and international trade. With some important exceptions, transport costs are typically treated as exogenous in the trade literature—approximated by distance empirically and by the "iceberg" functional form, where the value of shipped goods decreases ("melts") with distance. Trade also typically focuses on the value of trade flows, not on weight. However, while the collapse of the conventional measure of trade by value following the Great Recession is clearly visible in Figure 2, the measure of transportation usage by weight and distance barely changes. Even the costs of shipping actual icebergs are not well approximated by the "iceberg" functional form (Bosker and Buringh 2020). In contrast, in transportation economics the pricing structure is often at the per-unit level—for example, cost per ton or per container of goods. These transport prices/costs are equilibrium outcomes, jointly determined with trade and transport use. Both the assumptions of exogeneity and iceberg functional form, while providing tractability in most trade models, have nontrivial trade and welfare implications. We see bridging both literatures as a fruitful area for research.

Transport Costs over Time

How have the costs of global trade evolved? Data limitations make this question tricky to answer. While aggregate data on transportation expenditures are widely available, such data are rarely differentiated by whether the expenditures are for domestic or international trade. Additionally, while data on the value of internal trade flows are available for a subset of countries, data on distances covered internally are hard to come by—especially over our extended period.[3]

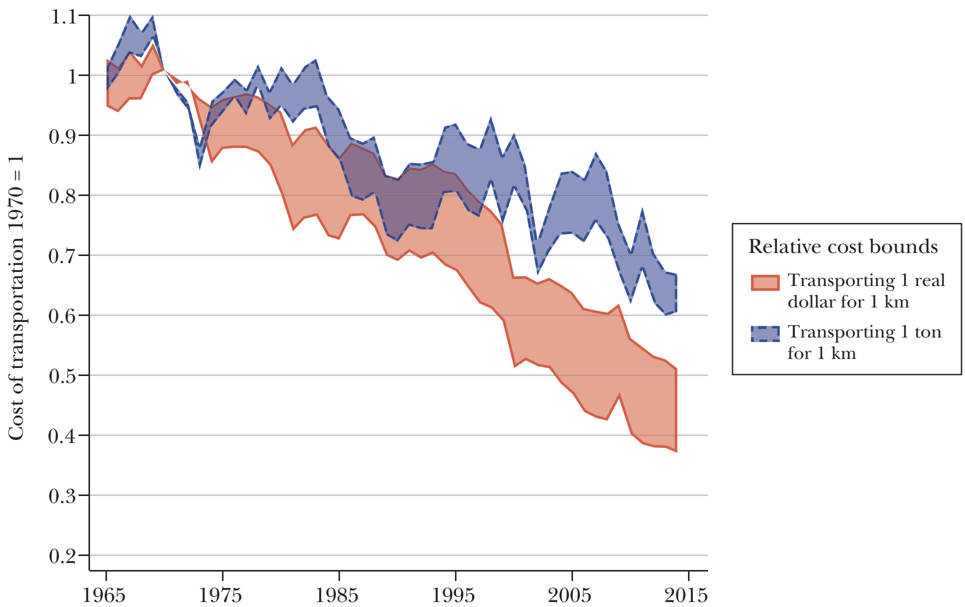
We describe a method of using total expenditures in the transportation sector—based on national accounting and aggregate industry data—to recover an upper and a lower bound for a price to ship either a ton or real dollar of goods for one kilometer. Our approach begins with the sum of all global transportation costs for a given year, divided by one of our measures of trade use for that year— either tons of trade or value of trade, multiplied by distance.

We first construct a cost estimate where all aggregate transportation spending was on international trade. Because we are dividing total transportation spending by international transport use, this approach effectively calculates an upper bound on the time trend of international trade costs.

For our lower-bound estimate, we include both international and domestic transport. We approximate domestic transport by assuming that the internal distance transported is unchanged over time. We consider this a lower bound for international trade costs, based on the assumption that while internal trade distances may have increased, these distances may increase at a slower rate than international trade distances. For this assumption, our sample is restricted to 24 countries with complete data over our time period from the WIOD (Timmer et al. 2015, see online Appendix for details and caveats). For our measure of the distance of internal domestic trade, we use the population-weighed, as-the-crow-flies internal distance between jurisdictions (like US states) from the CEPII gravity database, multiplied by the gross value of internal trade.

[3] While domestic transport and distribution costs are not our focus here, they can be nontrivial. Anderson and Van Wincoop (2004) estimate that domestic distribution cost can be 55 percent of producer prices, more than twice the international transport costs (echoed in India by Van Leemput 2021). Atkin and Donaldson (2015) show that intranational trade costs can be especially high in developing countries: four times larger in Nigeria than in the United States.

Figure 3
Transport Costs, 1965–2014



Source: BACI, WIOD, UN-NBER-Comtrade, PWT, and associated output deflators.
Note: Figure 3 is calculated using the sum of all global transportation costs for a given year, divided by trade use for that year—either tons of trade or value of trade, multiplied by distance (see equation [3] in online Appendix A for more details). For transportation costs, we use data on transportation and storage expenditures from the WIOD (ISIC codes I60–I63 [rev 3] and H49–H52 [rev 4]). The upper bound estimate is based on the scenario where all aggregate transportation spending was on international trade while the lower bound estimate reflects spending on both international and domestic trade. Values are normalized to one in 1970. Consistent sample of 24 countries representing 90 percent of world GDP. The time period for the WIOD ends at 2014 which restricts our sample (Timmer et al. 2015).

Our estimates, illustrated by Figure 3, show that show that global transport costs have substantially decreased from 1965 to 2014, reflecting large productivity increases and technological advances. The aggregate weight-based measure of transportation costs from 1970 to 2014 has fallen by 33–39 percent. Value-based measures have fallen by 48–62 percent.

Although measures of transport costs of goods by value and weight both show a downward trend, they provide different perspectives. The cost to transport one ton of goods for one kilometer decreased by about 35 percent over this period (gray dotted lines, Figure 3). This trend exhibits significant volatility from 1975–1985, reflecting in large part increases in the price of oil due to OPEC supply restrictions. Additionally, the cost of transporting a dollar's worth of goods for one kilometer has decreased by over 50 percent (red solid lines, Figure 3). This declining trend for a dollar of good means that this cost decline does not just apply to bulky goods, but also to all transported goods. In short, the transport-cost trends with trade use measured in dollars decrease faster than when trade is measured by weights.

This finding is generally consistent with Hummels (2007) in this journal, who documents a dramatic decline in transportation costs from 1950 to 2007 using direct data on prices paid for a consistent set of transportation modes, and highlights differences in quality and the endogenous selection of different modes of transport. It also lines up with Harrigan (2010) which finds that cheaper airfreight and containerization allows for the shipment of higher-value goods, while ocean bulk freight rates show less price movement.

Given that both transport costs have fallen and the global economy is using more transport services, one natural question to ask is whether aggregate spending on transport services has risen or fallen. We calculate global transport spending as a share of total gross output using the World Input-Output Database. The expenditure share on transport for these countries have mostly stayed constant: starting around 4 percent in 1965 and increasing to more than 4.7 percent by 1995, before declining back to 3.8 percent in 2015 (for details, see online Appendix Figure A1).[4]

The utilization of transport services is an equilibrium response to the cost of transport. One likely outcome is that these decreases in transport costs will alter the terms of the classic proximity-concentration trade-off: firms can either choose to expand production horizontally across borders to maximize proximity to foreign customers, or they can concentrate their production in one location in order to benefit from scale economies and export to these foreign destinations. There is little evidence that firms have sought to maximize proximity to foreign users.

Echoing our introductory case study, telephone manufacturers are no longer located close to their customers, and firms have generally expanded and even fragmented their production supply chains, altering the geographic location of economic activity (Antràs and Chor 2022; Redding 2022). Others have relied on lower transportation costs to expand exports; for example, using data on US multinational corporations, Brainard (1997) finds that as transport costs decrease, multinationals' exports outstrip overseas production. Helpman, Melitz, and Yeaple (2004) find the same empirical pattern in less developed countries.[5]

Finally, it is worth noting that transportation costs are not the only costs involved in trade. Indeed, direct transportation costs are a limited portion of overall distribution costs (Anderson and Van Wincoop 2004; Burstein, Neves, and Rebelo 2003), with wholesaling (Ganapati 2018; Chatterjee 2019) and retailing margins up to 50 percent. There are also costs of holding inventories and costs of time spent

[4] Using theory-based measures of aggregate trade costs that include all frictions to trade (as opposed to our accounting-based measure), Novy (2013) finds that US trade costs declined by about 40 percent between 1970 and 2000 while Jacks, Meissner, and Novy (2008) find a decrease of 16 percent from 1950-2000. In value-added terms, Redding and Turner (2015) show that US domestic transportation fell since 1965 from 4 percent to 3 percent of GDP. However, due to the nature of global shipping, this may be misattributed—the US economy may have simply outsourced some of its transportation costs.

[5] Bernard et al. (2020) find firms offshore only low quality products, that is, products with lower price-to-weight ratios (this is echoed in Lashkaripour 2020).

in the transportation process. When Head and Mayer (2013) attempt to line up trade costs with transportation and freight costs with the costs implied in standard trade models, they find that 50–90 percent of trade costs are generally unobserved. We will return to some of these issues later in discussions of the air and container freight, along with trade facilitation and inventory-holding behavior.

What Main Factors Have Contributed to Rising Transport Use?

What factors have contributed to the dramatic increases in transportation usage since 1965 as costs fell? We focus on three: (1) increasing participation of emerging economies, particularly China's, in global trade; (2) increasing trade between countries that are further apart; and (3) shifts in the composition of traded goods—natural resources, which are more upstream in supply chains, versus manufactured goods, which are more downstream.

Rise of Emerging Economies and China

Developing countries have increased their participation in world trade, accounting for about 40 percent of world exports in 2020 (UNCTAD 2022), but as one might expect, China is one of the driving forces behind this change.

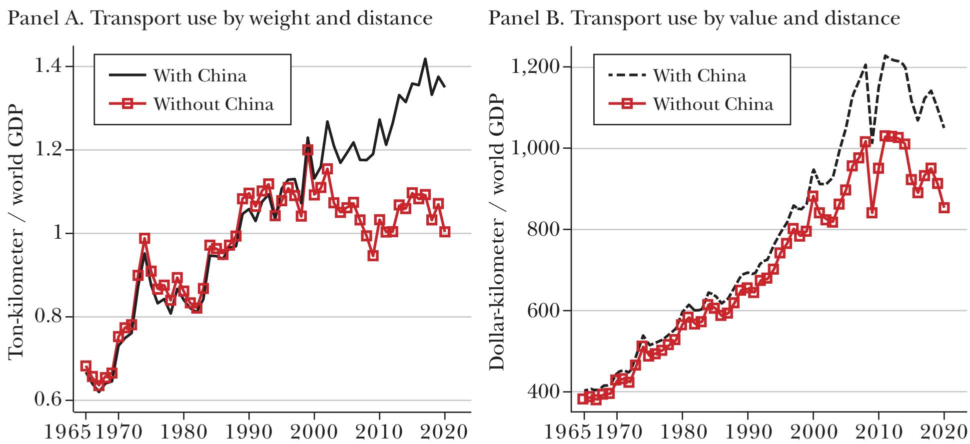
We recomputed our earlier measures of transportation use by weight and distance and by value and distance. For purposes of a rough comparison, we simply exclude both incoming and outgoing trade with China in the numerator as well as excluding Chinese GDP from world GDP in the denominator (see notes under Figure 4 for details). This metric is akin to considering world trade without China and not trying to model the many other consequences to international trade that would surely occur. We just assume that trade with China vanishes, along with Chinese GDP vanishing.[6]

We first consider the role of China in real transportation usage considering the ton-kilometers of goods (Figure 4, panel A). Between 1965 and the 1990s, transport usage with and without China is relatively similar. In 1965, \$1 of real output represents 0.67 ton-kilometers of transportation usage and 0.68 ton-kilometers without China. By 1990, both the global average and the average without China, rose to 1.06 and 1.10 ton-kilometers respectively for \$1 of output. At this point the two trends diverge. By 2020, the global average with China had increased by 28 percent since 1990, but the average excluding China had instead fallen by 9 percent. By the weight of goods, the growth of post-1990 transportation usage largely reflects the growth of China—to the near exclusion of many other trends.

Considering the value-based measure of trade, the dollar-distance trends with and without China mostly increase in tandem (Figure 4, panel B), although we do see a divergence starting in 1990 that is much smaller and slower than

[6]A fuller experiment would be to embed China in a model that simultaneously computes both trade flows as well as transportation usages, as in Ganapati, Wong, and Ziv (2021).

Figure 4
The Role of Emerging Economies, Focusing on China, 1965–2020



Source: BACI, WIOD, UN-NBER-Comtrade, PWT, and associated output deflators.
Notes: The real transport use measured in ton-distance, indicated by the black line in panel A, is reproduced from Figure 1, panel C, for comparison. The real transport use measured in ton-distance excluding China, indicated by the red line with squares in panel A, is calculated by recomputing equation (2) in online Appendix A to exclude both incoming and outgoing trade in weight with China in the numerator, as well as excluding Chinese GDP from world GDP in the denominator. The real transport use measured in dollar-distance, indicated by the black line in panel B, is reproduced from Figure 1 panel D for comparison. The real transport use measured in dollar-distance excluding China, indicated by the red line with squares in panel A is calculated by recomputing equation (2) in online Appendix A to exclude both incoming and outgoing trade in value with China in the numerator, as well as excluding Chinese GDP from world GDP in the denominator. See Figure 1 notes and the online Data Appendix for further details.

the weight-distance trend. By 2007, the distance traveled for $1 of final good consumption was 1,165 dollar-kilometers, while the number without China was 976 dollar-kilometers. Following the great recession, both numbers, with and without China, faced a similar decline to 1,049 dollar-kilometers and 853 dollar-kilometers respectively.

China's growth trajectory offers one reason for these differential trends. In 1990, China was importing and exporting raw materials and other similar low-value, high-weight products. These trends continued through the Great Recession, especially to and from China. However, China also started exporting and importing high-value, low-weight goods, including large volumes of electronic components and smartphones. Trade in these goods, in China as in the rest of the world, leveled off following 2007. The collapse in value-based measures of trade after 2007 is not a China-specific trend (Baldwin and Evenett 2009; Eaton et al. 2016). However, these high-value, low-weight goods may face changes in the proximity-concentration trade-off. As it becomes cheaper for those raw materials to be imported, production locations for these products may be more reflective of where final demand is located.[7]

Longer-Distance Trade

Most of the growth in real transport usage in terms of weight is due to a rise in longer-distance trade between countries that are further apart. Figure 5 breaks down our two metrics for real transportation usage from Figure 1, panels C and D, into three sub-components: shorter distance trade under 5,000 kilometers, medium distance trade from 5,000 to 10,000 kilometers, and long distance trade over 10,000 kilometers. The short-distance bin, within 5,000 kilometers, typically includes country pairs that are in the same region (for example, countries within the East Asian region, the European Union, or North America). The medium-distance bin (5,000-10,000 kilometers) typically includes Asian and European countries, while the long-distance bin includes Asian and North American countries. The transport use of all three distance bins adds up to the aggregate transport use measure previously presented in Figure 1.

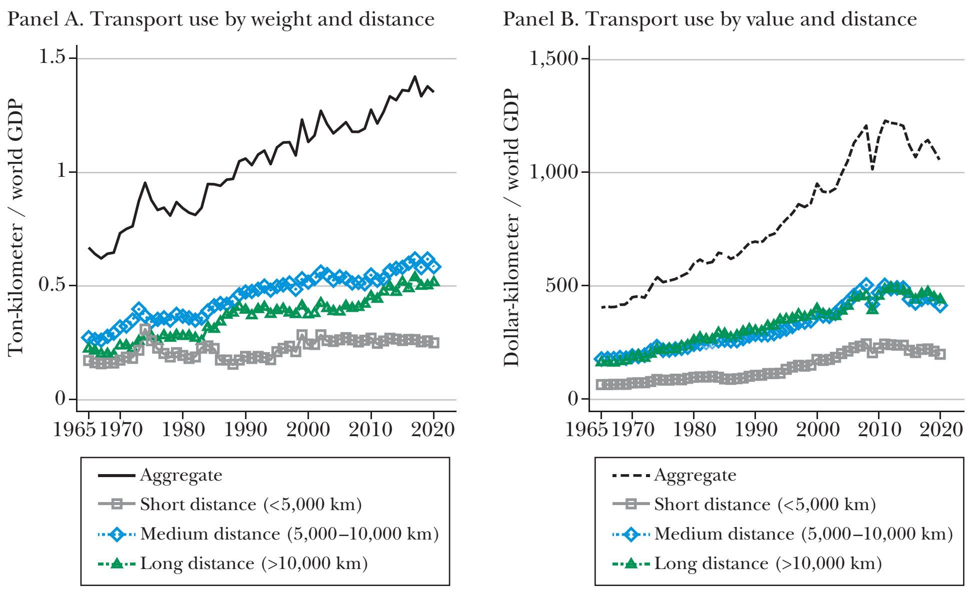
In 1965, all three distance bins account for roughly similar amounts of transport usage by weight (Figure 5, panel A). But since the mid-1980s, the transport use by countries that are further apart increases by much more than the short-distance countries. Overall, short-distance countries increase their transport use in weight by 45 percent from 1965 to 2020, while longer-distance countries more than doubled their transport use—medium-distance countries increased by 114 percent and long-distance countries by 129 percent. Digging into the underlying data, much of the increase stems from raw material shipments, often originating from OPEC countries, Australia, and Brazil. These raw materials are often bound for processing and usage in distant locations, especially China.

When considering the transportation usage in value, Figure 5, panel B, highlights a different story. There are large increases across all three distances. Shipments for short distances increased by 211 percent, medium distances by 134 percent, and long distances by 170 percent. The rough similarity of these trends highlights the near identical growth paths of trade value and transport use by value in Figure 2 over our period.

Overall, while heavier and lower value goods are being traded between countries that are further apart, lighter and higher value goods are being traded between locations that are both nearby and far apart. To give an intuitive example, while the raw materials to make phone components are traveling even farther than before, so are the final smartphones themselves.

[7] Antràs et al. (2012) finds that emerging economies play different roles in global supply chains. Bangladesh is one of the most downstream countries in terms of manufacturing because it exports apparel that are sold directly to end consumers, while Tajikistan is one of the most upstream countries because it exports processed alumina.

Figure 5
The Role of Longer Distance Trade, 1965–2020



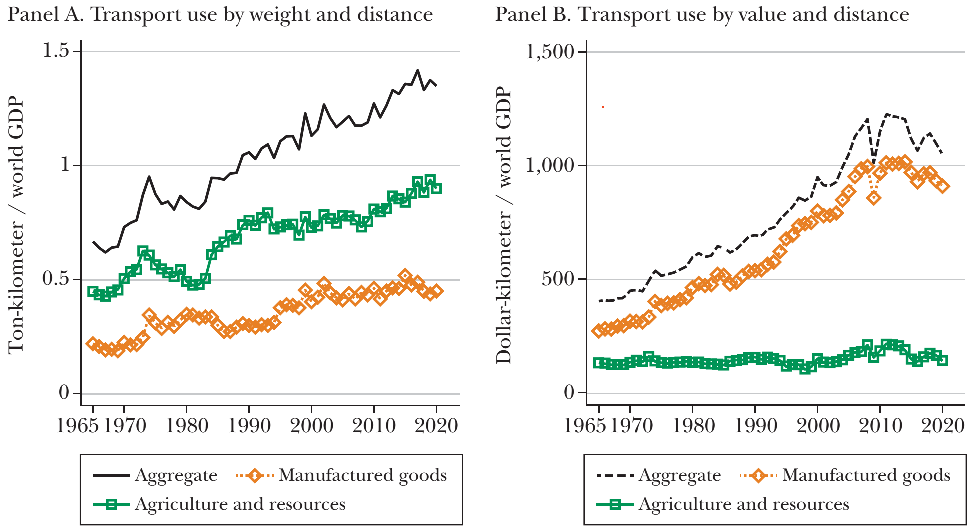
Source: BACI, WIOD, UN-NBER-Comtrade, PWT, and associated output deflators.
Notes: The real transport use measured in ton-distance, indicated by the black line in panel A, is reproduced from Figure 1, panel C, for comparison. The remaining three lines in panel A are calculated by breaking down the real transport use measure in weight into three subcomponents that total the aggregate figures: transportation usage of shorter-distance trade under 5,000 kilometers (gray line with squares), medium-distance trade from 5,000 to 10,000 kilometers (blue line with diamonds), and long-distance trade over 10,000 kilometers (green line with triangles). The real transport use measured in dollar-distance, indicated by the black line in panel B, is reproduced from Figure 1, panel D, for comparison. The remaining three lines in panel B are calculated by breaking down the real transport use measure in value into three subcomponents that total the aggregate figures: transportation usage of shorter-distance trade under 5,000 kilometers (gray line with squares), medium-distance trade from 5,000 to 10,000 kilometers (blue line with diamonds), and long-distance trade over 10,000 kilometers (green line with triangles). See Figure 1 notes and the online Data Appendix for further details.

Composition of Trade

What types of goods have contributed most to the rise of transportation usage over the past 50 years—and where are these goods on the value chain? We study the how the composition of trade has contributed to transport use over this period by highlighting the role of raw materials—that is, agricultural and natural resource products—relative to manufactured goods in Figure 6.

Figure 6, panel A, shows that while raw materials make up a higher share of aggregate transport use by weight throughout this period (accounting for 66 percent of the weight-distance measure in 2020), both raw materials and manufactured goods equally contributed to the growth from 1965 to the early 2000s. Since 2000,

Figure 6
The Role of Natural Resources versus Manufactured Goods, 1965–2020



Source: BACI, WIOD, UN-NBER-Comtrade, PWT, and associated output deflators.
Notes: The real transport use measured in ton-distance, indicated by the solid black line in panel A, is reproduced from Figure 1, panel C, for comparison. The remaining two lines in panel A are calculated by breaking down the real transport use measure in weight into two subcomponents that total the aggregate figure: agricultural and natural resource products (ISIC categories A and B, green line with squares), and manufactured products (orange line with diamonds). The real transport use measured in dollar-distance, indicated by the dashed black line in panel B, is reproduced from Figure 1, panel D, for comparison. The remaining two lines in panel B are calculated by breaking down the real transport use measure in value into two subcomponents that total the aggregate figure: agricultural and natural resource products (ISIC categories A and B, green line with squares), and manufactured products (orange line with diamonds). See Figure 1 notes and the online Data Appendix for further details.

however, manufactured goods no longer contribute to the growth in ton-kilometers. Instead, all growth is due to raw materials.

Revisiting our phone example, raw materials have been essential since the early days of telephone manufacturing. Even though all US telephones were essentially manufactured in a single Western Electric factory for decades, the manufacturing process still required 34 different raw materials including metal ores, rubber, mica, silk, and cotton from countries including India, Indonesia, Brazil, Madagascar, Japan, and China (Western Electric 1938). These raw materials play a central role in manufacturing and have relatively high weight-to-value ratios—contributing more to weight-based measures of transport use.

Considering the value-distance of shipped goods in Figure 6, panel B, we find very different results: the increase in transport use by value over this period are entirely driven by manufactured goods. While manufactured-goods transport use grew 135 percent, that of raw materials grew by just 6 percent. Even though raw materials are a required input in much of economic production, as reflected by their dominance of transportation usage in ton-kilometers, they constitute a smaller and smaller share of total trade values. An assembled smartphone, even including packaging, may be under 250 grams, taking up less than a liter of volume, and has high value-to-weight ratio. The raw materials that are used in smartphone manufacturing, even though crucial, make up a small share of its final value.[8]

Our finding that raw materials dominate the transport use growth by weight post-2000 (and in levels), but not value, is consistent with Fally and Sayre (2018). While primary commodities—intensive in natural resources—only account for a modest 16 percent of world trade by value, these commodities are used as inputs into all production processes, are difficult to find substitutes for, and can be supplied by only a few countries. As the world economy grows, these raw materials are shipped farther distances (echoed in Berthelon and Freund 2008). As trade costs fall, Antràs and de Gortari (2020) show that locating downstream production close to consumers is less important than for upstream production.

Innovations and Implications for Supply Chains

We have established two main results so far. First, transport use has increased while transport costs have decreased over the past 50 years. Second, this increased demand for transportation is driven by emerging economies—especially China—participating in world trade, longer-distance trade between countries that are further apart, and differences in where products lie along the value chains. In this section, we consider the main drivers of changes in these international transport costs over the last five decades and their implications up and down supply chains.

We first highlight innovations in transport technology and infrastructure, especially in container and air shipping. We then outline developments in trade networks, trade facilitation, and investments in infrastructure. For an overview of the extensive literature that indirectly looks at the impacts of transport costs, we recommend Redding and Turner (2015).

Transport Technology Innovations: Container and Air Freight Shipping

Global trade is conducted using land, sea, and air. Transport costs have fallen across all three modes in the last 50 years (Ardelean et al. 2022). Two technologies have exhibited extraordinary cost decreases: containerized and air freight. Hummels (2007) documents that the cost of air transport fell by a factor of more than ten between 1955 and 2004, and the container price index declined rapidly between 1985 and 2004. We focus on these innovations here.

[8] In online Appendix A, we replicate this analysis with final and intermediate goods and find qualitatively similar results (online Appendix Figure A3).

Containerization refers to the standardization of a 40-foot-long reusable steel box that can be loaded onto purpose-built trains, trucks, and ships, easily transferring between various modes from origin to destination. After its introduction in 1956, almost 90 percent of countries had container-handling infrastructure by 1983 (Rua 2014), making the globalization of production possible (UNCTAD 2022).

Container ships offer a useful example of natural scale economies. The size of container ships has increased significantly over the years (Cullinane and Khanna 2000). Early container ships in the 1960s were modified bulk vessels, with capacities of 500 TEUs (as we mentioned earlier, this stands for twenty-foot equivalent units). By the start of the 1970s, the first ships dedicated to transporting containers were introduced with two to four times the capacities of their predecessors (Rodrigue 2020). By the end of our sample period in 2015, the largest container ship built that year (the *MSC Oscar*) had a capacity 24 times the size of the first containerships—19,000 TEUs and nearly as long as four football fields laid end-to-end. To put the capacity of this ship in perspective, it can carry 39,000 cars, 117 million pairs of sneakers, or more than 900 million cans of dog food (Stromberg 2015).

While the total capacity of the containerized fleet increased eight times from 1996 to 2021, the number of containerships only increased by three-fold (Ardelean et al. 2022). As crew size and fuel costs do not increase proportionally with ship sizes, larger ships take advantage of the cost savings from larger capacities. In 2013, the chief executive officer of the shipping firm Maersk explained that their decision to build larger ships was due to saving $300 to $400 per container for a round trip between Asia and Europe—that is, a savings of $5–$7 million per trip (Milne 2013).[9]

Time can also be a highly valued characteristic of trade (Hummels and Schaur 2013), and reliable and frequent air transport networks have made it much easier to coordinate the production of sensitive products across global transport chains. Increased air connectivity has positive impacts on local economic activity including population and income (Blonigen and Cristea 2015), industrial activity (Redding, Sturm, and Wolf 2011), and business links (Campante and Yanagizawa-Drott 2018). Additionally, it can have indirect positive effects on trade (Cristea 2011; Poole 2016; Yilmazkuday and Yilmazkuday 2017). For a useful overview of the development of air freight, see Proctor, Machat, and Kodera (2010).

What kinds of goods are more likely to be transported by containers and air, and can we characterize where are they on the supply chain? The use of containerization by land- and sea-based transportation accounts for many of the differences in trends between weight-based and distance-based measures. Moreover, the nature of containerization implies high fixed costs and minimal marginal costs, as highlighted by Coşar and Demir (2018). Because container cost is typically per-unit, Hummels

[9] Of course, the extent to which scale economies are passed along to buyers will be affected by the extent of completion. In transportation, Hummels, Lugovskyy, and Skiba (2009) and Asturias (2020) show large effects of market entry on reducing prices in international shipping. Ignatenko (2021) and Ardelean and Lugovskyy (2023) considers the effects of competition on the extent of price discrimination.

and Skiba (2004) show that per-unit transactions cost lowers the relative price of, and raises the relative demand for, high-quality goods with higher unit values.

A general pattern arises among air, containers, and bulk shipments: goods that travel by air have the highest value-to-weight ratio, followed by containers, and then bulk shipments.[10] In the global shipping fleet, for example, ships that only carry bulk goods, including agriculture, natural resources, and refined petroleum products, account for over 75 percent of the global shipping fleet by tonnage. Container ships—the ships that carry significant amount of consumer goods—only account for 13 percent. In terms of tons loaded, containers account for less than 2 million tons loaded—out of more than 10 million tons (UNCTAD 2021).

As goods move down the supply chain, their value-to-weight ratio rapidly changes. In our earlier example, a smartphone such as an Apple iPhone may require tons of coal and oil to generate electricity, bauxite to create aluminum enclosures, lithium ore for batteries, and sand to make silicon. These raw materials have very low value-to-weight ratios and are transported by means of bulk shipping. However, the assembled smartphone has a much higher value-to-weight ratio and is often transported by air.

Overall, the technological improvements from containerization and air transport have revolutionized the shipment of high-value final goods and downstream manufactured products, but have had minimal impact on the trade of raw materials and upstream manufactured products which rely heavily on bulk transport. Recalling our earlier discussion (in Figure 3) that trade costs using dollars have fallen between 48 and 62 percent over the relevant time period, while trade costs using physical weight has only fallen 33–39 percent. Thus, trade costs have fallen more for higher-value goods far down the value chain and less for goods further upstream. Correspondingly, transport use from downstream manufactured goods (compared to raw materials) contributed disproportionally to transport use growth by value (Figure 6, panel B).

For raw materials and bulk shipments, the corresponding change is not a greater quantity shipped, but a greater distance shipped. Figure 6, panel A, further reinforces this point by showing that raw materials make up a higher aggregate share of transport use by weight over the past 50 years and contribute to all of the increase in transport use by weight in the last two decades.

Endogenous Transport Costs

Transport costs are equilibrium outcomes jointly determined with trade and transport. Interdependencies in transportation technologies create networks and feedback loops which can magnify reductions in cost of the underlying technological improvements to transportation. Improving access within trade networks, like the 2016 Panama Canal expansion, can generate multiplicative trade returns (Heiland et al. 2019). We highlight two examples of how transport costs can be endogenous

[10] We omit discussion of pipelines versus tankers. Pipelines offer cheaper shipping, but are only built in response to high expected usage.

and their implications for international trade: (1) country linkages within round-trip routes, and (2) hub-and-spoke network effects and scale economies.

When modern container technology was introduced in 1956, it triggered complementary technological and logistical innovations that have revolutionized the transport industry and international trade (Bernhofen, El-Sahli, and Kneller 2016; Levinson 2016). Examples of complementary innovations include shipping capacity increases through larger ships, automated port infrastructure, and delivery-time reductions through unified logistic communication systems. Containerization also facilitated multimodal transportation—it is much easier to have a crane move a container from ship to rail or truck at ports (Fuchs and Wong 2022).

These innovations result in linkages between countries from round-tripping: container ships, trucks, and air transport have fixed schedules, like buses, going back and forth between large trading partners in a round trip. If trade of goods is flowing mostly in one direction, the "backhaul problem" arises of how to make use of transportation capacity for its return journey (Jonkeren et al. 2011; Tanaka and Tsubota 2016; Friedt and Wilson 2020) and can shape the location of economic activity in the presence of agglomeration (Behrens and Picard 2011). These back-and-forth dynamics can also create backfiring effects from protectionist policies when limits on trade in one direction lead to less capacity for trade flowing in the other direction (Hayakawa, Ishikawa, and Tarui 2020; Wong 2022).

The trade network of container (and air) shipping is a hub-and-spoke system where majority of trade is shipped indirectly—the median shipment to the United States stops at two additional countries before its destination (Ganapati, Wong, and Ziv 2021). The majority of these additional countries are hubs, or entrepôts, that play important roles in consolidating goods from nearby countries into larger ships, taking advantage of scale economies, and also connecting countries to each other globally by allowing countries to ship indirectly by means of the network. Similar phenomena occur elsewhere. Bulk shippers transport goods directly, but often have to search for loading opportunities after delivering their cargo, generating network effects between neighboring countries (Brancaccio, Kalouptsidi, and Papageorgiou 2020).

Emerging economies like China and longer-distance trade have driven the increase in transport use in recent decades (shown in Figures 4 and 5). Indirect trade by means of trade networks facilitates this increase—while larger developed and emerging countries can utilize larger ships due to their size, smaller countries that are more remote are also able to take advantage of the same scale economies from larger ships when their goods are routed through entrepôts (Ganapati, Wong, and Ziv 2021). The country-level connections within a round-trip further contributes to these linkages.

Trade Facilitation and Infrastructure Improvements

"Trade facilitation" refers to policies that lower administrative barriers to trade by streamlining administrative processes, like filing of shipment documents at border crossings, which in turn decreases the management cost of supply chains (for an overview, see Carballo, Schaur, and Martincus 2018).

Various estimates suggest that at least half of trade costs are not observed in the aggregate national statistics of transportation spending (Head and Mayer 2013; Feyrer 2021; Allen 2014). However, the logistics and transportation industry have no trouble naming costs that economists label as “missing.” We highlight two such costs here: logistic management technology and the services of freight forwarders.

Better computing technology and efficiency in logistics allows companies to coordinate large volumes of shipments to different locations, port authorities to manage these shipments through their ports, and shipping carriers to keep better track of containers on their ships. One prominent example is the introduction of the cargo-booking documents system called INTTRA in the early 2000s. It allows nonvessel-owning owning carriers and freight forwarders (discussed in a moment) to book cargo and access voyage schedules. Another example is the introduction of the Society for Worldwide Interbank Financial Telecommunication (SWIFT) messaging system in the 1970s, which allows for efficient and secure transfer of funds by banks between importers, exporters, and transportation intermediaries, lowering their financial transaction costs. For an overview of the logistics of international trade, see Talley and Riggs (2018) or Hesse and Rodrigue (2004).

Transportation intermediaries like freight forwarders are responsible for cargo pickup, documentation, transport, and delivery from the beginning to the end of the value chain (for an overview, see Blum, Claro, and Horstmann 2018). Without these middlemen, a trader would have to coordinate multiple separate steps: transport from factory to port, ship to a destination port, clearing customs, and then transport from port to the final destination. With a freight forwarder, the exporter only has to interface with one company. Container technology has contributed to the growth of freight forwarders, and large-scale services were offered starting in the early 1970s (UNCTAD 2021). By 2018, major container shipping lines who only provide port-to-port service have largely disappeared from the market. Such middlemen roles are not new, and in other settings they are instrumental in facilitating trade (Ganapati 2018; Grant and Startz 2022).

Improved tools for logistics, transportation intermediaries, and national policies to facilitate border crossings all work together to reduce the broader transportation costs experienced in international trade. As transport use continues to rise and trade between countries that are further apart continues to grow, investments in transportation infrastructure play an increasingly important role; indeed, the quality of transport infrastructure has been shown to be directly proportional to transport costs (Limao and Venables 2001). Port efficiency can play a major role in facilitating trade flows (for an overview, see Blonigen and Wilson 2018). Other infrastructure investments, including railroad networks, pipelines, and high-capacity expressways, can have large benefits and result in direct decreases on transport costs (Coşar and Demir 2016; Donaldson 2018; Fan, Lu, and Luo 2019). In the case of emerging economies, infrastructure improvements could have even larger welfare impacts (Asturias, García-Santana, and Ramos 2019; Bonadio 2021; Carballo et al. 2021).

Just-in-Time Production

The less expensive and more reliable international transportation linkages have helped to spawn greater use of just-in-time deliveries, also called "lean manufacturing," in which inputs are received from suppliers only as needed for the production process. The benefits claimed for just-in-time deliveries include smaller inventories, less waste, and continual two-way feedback between suppliers and buyers. Toyota offers most famous and well documented case, which significantly reduced warehouse and inventory costs (Moore 2010). This broader shift in business practice is also visible in US Census Bureau and Bureau of Economic of Analysis data. For example, in US manufacturing, the inventory-to-sales ratio was around 1.7 before 1990. Since then, there has been steady decrease in the amount of inventory that businesses hold—the average ratio is about 1.3 between 2000 and 2019.[11]

Going Forward: Resiliency and Vulnerabilities

In the aftermath of the supply disruptions during the COVID-19 pandemic, a key question is whether the patterns and changes we have described in this paper made modern supply chains and networks more resilient or more vulnerable to shocks. This is a growing area of research. Such shocks may be related to trade policy, like Brexit or the 2022 trade sanctions on Russia, or may relate to events that directly affect production and transportation, whether from pandemic or other natural disasters. Examples of such research include Boehm, Flaaen, and Pandalai-Nayar's (2019) consideration of how shocks from earthquakes propagate into upstream production; Feyrer's (2021) study of the Suez Canal closure on aggregate trade flows; Khanna, Morales, and Pandalai-Nayar's (2022) look at manufacturers during COVID-19 lockdowns and highlighting of the role of multiple sourcing; and Besedeš and Murshid's (2019) study of an Icelandic volcano eruption that closed European airspace. This is a nascent literature, and here are some themes and connections we would emphasize as this research develops.

First, goods in the world economy are being transported over increasingly longer distances, by more diverse sets of countries, and often traveling indirectly to their destination, which further lengthens their trips. Longer shipment distances are mechanically more vulnerable to transportation disruptions because potential shocks can affect more locations. Similarly, longer-distance trade may have to cross many more choke points, like the Suez and Panama Canals and the Straits of Malacca and Hormuz. The obstruction of the Suez Canal for six days in March 2021, and its subsequent supply chain disruptions, serves as an illustrative example of how longer-distance trade can be more subjected to these potential disruptions. Trade may be less resilient to shocks because of the increasing importance of such chokepoints, but little research has touched on long run trends here and little is known concretely.

[11] For additional details on patterns of US inventories, see online Appendix Figures A4 and A5.

Second, transportation networks are not only longer today, but also feature round-trip and hub-and-spoke networks that operate on fixed routes. With fixed-schedule routes, a disruption to one leg of a trip not only affects goods on that leg, but can cascade throughout the network (Swanson 2021). Additionally, larger and larger ships—both container ships and bulk freight—are built to take advantage of the per unit cost savings. These large ships concentrate the hub and spoke system further, and utilize the multimodal transport network more, resulting in further international links between countries which spill over into domestic intranational links between cities or regions as well. Changes to relationships between countries, like the recent disturbing turn to protectionism in parts of the world, have much more widespread spillover impacts due to transport networks and supply chains, and deserve further research.

Third, some vulnerabilities to trade may not be readily apparent. For example, when production of final goods happens closer to home, it may appear less risky due to less exposure to trade. However, this local production still requires upstream inputs in their production, and their ultimate reliance on trade and transport requires analysis of the entire value chain and its alternatives. For example, while air transport may be considered to have a lower risk in terms of time delays compared to ocean shipping through the Suez Canal, air goods are produced using goods that are shipped by ocean—and thus may be exposed to ocean disruptions as well.

Fourth, the "bullwhip" effect refers to a situation where small perturbations upstream in a production chain are amplified downstream and become major issues (Fransoo and Wouters 2000; Lee, Padmanabhan, and Whang 1997). Upstream production may use input products with a limited set of globally dispersed substitutes, while downstream may face many different alternatives that are more geographically concentrated. Several issues already mentioned, including longer distances, networks, choke points, congestion, and scale economies, can be combined to create a bullwhip effect.[12]

With these issues in mind, we highlight two additional areas for fruitful research: the nexus between the environmental impacts of transportation and supply chains, and the interaction between market power and long run trends in transportation networks.

While transportation only accounts for 15 percent of greenhouse gas emissions, it also enables the shifting of pollution in other sectors across space, especially in agriculture and industry (Shapiro 2016). While a small literature considers how environmental policy can have adverse unintended effects on overall pollution within transportation (Cristea et al. 2013; Mundaca, Strand, and Young 2021; Lugoskyy, Skiba, and Terner 2022), few studies consider the role of transportation

[12] Some papers have considered the endogenous responses of transportation costs to shocks (Fajgelbaum and Schaal 2020; Allen and Arkolakis 2022; Ganapati, Wong, and Ziv 2021; Brancaccio, Kalouptsidi, and Papageorgiou 2020). However, these studies are limited to looking at just fragments of both the transportation network and the value chain; none of these papers can integrate the upstream and downstream effects across modes of transport.

costs on both the distribution of pollution as well as on the aggregate levels though their interaction with other sectors.

Second, larger markets induce entry, driving down markups and prices, even if costs are constant. As absolute demand for transportation services increases, entry can further amplify the affect of entrants helping to discipline costs. In transportation, Hummels, Lugovskyy, and Skiba (2009) show large effects of market entry on reducing prices in international shipping. However, a second countervailing trend exists, that of scale economies: as ships get bigger and airline shipping networks get denser, a smaller and smaller set of firms may dominate an industry and potentially extract large profits—muting the gains from trade.

Although the vulnerability of global supply chains is at the forefront of many minds (including our own!) in the aftermath of the post-pandemic congestion and disruptions, the long-term trends in this paper suggest a more nuanced approach to these issues. If one compares trade between, say, the United States or the European Union and a wide range of destinations around the world, for most of those destinations trade is considerably more resilient and less vulnerable than several decades ago—and those gains to resiliency are in fact apparent in the greater use of trade over longer distances, the expanded trade networks, better payment systems, better trade facilitation, and so on. Countries, and cities within countries, are more cohesively and reliably interconnected than in earlier decades because of the interaction between efficient transport and supply chains.

In an earlier example of supply chain vulnerability back in February 1997, a fire occurred at a Toyota parts supplier that was the sole supplier of a crucial part all Toyota vehicles. The just-in-time inventory system meant that the resulting ripple effect shut down all Toyota production for two weeks. But Toyota and other firms interviewed by Nishiguchi and Beaudet (1998) did not consider abandoning just-in-time, but instead focused on developing flexibility within their firm to better respond to future issues like this. Similarly, when faced with the recent supply disruptions and congestion of the pandemic, firms and countries should not start looking inwards in terms of production processes and decrease their transport use. Instead, unexpected events teach lessons about unforeseen risks; for example, it is not only important for a firm to have multiple suppliers, but also to know that those suppliers can use multiple trade routes. Consolidating the entire supply chain at home is both costly and as risky, albeit in a different way, as a single-sourced foreign location. Researchers should be able to study both sides of the coin: on one side, the gains to the world economy from increased transport reliability and use of trade, and on the other side, how local disruptions to production can have far and wide-reaching consequences through the interactions of trading networks.

■ *We thank our editors Erik Hurst, Nina Pavcnik, Timothy Taylor, and Heidi Williams for extremely helpful feedback, as well as Mark Colas and David Hummels for constructive comments.*

References

Allen, Treb. 2014. "Information Frictions in Trade." *Econometrica* 82 (6): 2041–83.

Allen, Treb, and Costas Arkolakis. 2022. "The Welfare Effects of Transportation Infrastructure Improvements." *Review of Economic Studies* 89 (6): 2911–57.

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

Antràs, Pol, and Davin Chor. 2022. "Global Value Chains." In *Handbook of International Economics: International Trade*, Volume 5, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 297–376. Amsterdam: Elsevier.

Antràs, Pol, and Alonso de Gortari. 2020. "On the Geography of Global Value Chains." *Econometrica* 88 (4): 1553–98.

Antràs, Pol, Davin Chor, Thibault Fally, and Russell Hillberry. 2012. "Measuring the Upstreamness of Production and Trade Flows." *American Economic Review* 102 (3): 412-16.

Ardelean, Adina, Volodymyr Lugovskyy, Alexandre Skiba, and David Terner. 2022. "Fathoming Shipping Costs: An Exploration of Recent Literature, Data, and Patterns." World Bank Policy Research Working Paper 9992.

Ardelean, Adina, and Volodymyr Lugovskyy. 2023. "It Pays to Be Big: Price Discrimination in Maritime Shipping." *European Economic Review* 153: 104403.

Asturias, Jose. 2020. "Endogenous Transportation Costs." *European Economic Review* 123: 103366.

Asturias, Jose, Manuel García-Santana, and Roberto Ramos. 2019. "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India." *Journal of the European Economic Association* 17 (6): 1881–1940.

Atkin, David, and Dave Donaldson. 2015. "Who's Getting Globalized? The Size and Implications of Intranational Trade Costs." NBER Working Paper 21439.

Baldwin, Richard, and Simon Evenett, eds. 2009. *The Collapse of Global Trade, Murky Protectionism, and the Crisis: Recommendations for the G20.* London: Centre for Economic Policy Research.

Behrens, Kristian, and Pierre M. Picard. 2011. "Transportation, Freight Rates, and Economic Geography." *Journal of International Economics* 85 (2): 280–91.

Bernard, Andrew B., Teresa C. Fort, Valerie Smeets, and Frederic Warzynski. 2020. "Heterogeneous Globalization: Offshoring and Reorganization." NBER Working Paper 26854.

Bernhofen, Daniel M., Zouheir El-Sahli, and Richard Kneller. 2016. "Estimating the Effects of the Container Revolution on World Trade." *Journal of International Economics* 98: 36–50.

Berthelon, Matias, and Caroline Freund. 2008. "On the Conservation of Distance in International Trade." *Journal of International Economics* 75 (2): 310–20.

Besedeš, Tibor, and Antu Panini Murshid. 2019. "Experimenting with Ash: The Trade-Effects of Airspace Closures in the Aftermath of Eyjafjallajökull." Unpublished.

Blonigen, Bruce A., and Anca D. Cristea. 2015. "Air Service and Urban Growth: Evidence from a Quasinatural Policy Experiment." *Journal of Urban Economics* 86: 128–46.

Blonigen, Bruce A., and Wesley W. Wilson. 2018. "Trade Costs, Trade, and Port Efficiency." In *Handbook of International Trade and Transportation*, edited by Bruce A. Blonigen and Wesley W. Wilson, 478–99. Cheltenham, UK: Edward Elgar Publishing.

Blum, Bernardo S., Sebastian Claro, and Ignatius J. Horstmann. 2018. "Trade Costs and the Role of International Trade Intermediaries." In *Handbook of International Trade and Transportation*, edited by Bruce A. Bloningen and Wesley W. Wilson, 337–68. Cheltenham, UK: Edward Elgar Publishing.

Boehm, Christoph E., Aaron Flaaen, and Nitya Pandalai-Nayar. 2019. "Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tōhoku Earthquake." *Review of Economics and Statistics* 101 (1): 60–75.

Bonadio, Barthélémy. 2021. "Ports vs. Roads: Infrastructure, Market Access and Regional Outcomes." Unpublished.

Bosker, Maarten, and Eltjo Buringh. 2020. "Ice(berg) Transport Costs." *Economic Journal* 130 (629): 1262–87.

Brainard, S. Lael. 1997. "An Empirical Assessment of the Proximity-Concentration Trade-Off between Multinational Sales and Trade." *American Economic Review* 87 (4): 520–44.

Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou. 2020. "Geography, Transportation, and Endogenous Trade Costs." *Econometrica* 88 (2): 657–91.

Burstein, Ariel T., Joao C. Neves, and Sergio Rebelo. 2003. "Distribution Costs and Real Exchange Rate Dynamics during Exchange-Rate-Based Stabilizations." *Journal of Monetary Economics* 50 (6): 1189–1214.

Campante, Filipe, and David Yanagizawa-Drott. 2018. "Long-Range Growth: Economic Development in the Global Network of Air Links." *Quarterly Journal of Economics* 133 (3): 1395–1458.

Carballo, Jerónimo, Alejandro G. Graziano, Georg Schaur, and Christian Volpe Martincus. 2021. "The Effects of Transit Systems on International Trade." *Review of Economics and Statistics*. https://doi.org/10.1162/rest_a_01189.

Carballo, Jerónimo, Georg Schaur, and Christian Volpe Martincus. 2018. "Transportation and Trade Interactions: A Trade Facilitation Perspective." In *Handbook of International Trade and Transportation*, edited by Bruce A. Blonigen and Wesley W. Wilson, 422–50. Cheltenham, UK: Edward Elgar Publishing.

Chatterjee, Shoumitro. 2019. "Market Power and Spatial Competition in Rural India." Cambridge Working Paper in Economics 1921.

Conte, Maddalena, Pierre Cotterlaz, and Thierry Mayer. 2021. "Data for: The CEPII Gravity Database." CEPII Working Paper 2022-05.

Coşar, A. Kerem, and Banu Demir. 2016. "Domestic Road Infrastructure and International Trade: Evidence from Turkey." *Journal of Development Economics* 118: 232–44.

Coşar, A. Kerem, and Banu Demir. 2018. "Shipping inside the Box: Containerization and Trade." *Journal of International Economics* 114: 331–45.

Cristea, Anca D. 2011. "Buyer-Seller Relationships in International Trade: Evidence from U.S. States' Exports and Business-Class Travel." *Journal of International Economics* 84 (2): 207–20.

Cristea, Anca, David Hummels, Laura Puzzello, and Misak Avetisyan. 2013. "Trade and the Greenhouse Gas Emissions from International Freight Transport." *Journal of Environmental Economics and Management* 65 (1): 153–73.

Cullinane, Kevin, and Mahim Khanna. 2000. "Economies of Scale in Large Containerships: Optimal Size and Geographical Implications." *Journal of Transport Geography* 8 (3): 181–95.

Dedrick, Jason, and Kenneth L. Kraemer. 2017. "Intangible Assets and Value Capture in Global Value Chains: The Smartphone Industry." World Intellectual Property Organization Economic Research Working Paper 41.

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review* 108 (4–5): 899–934.

Eaton, Jonathan, Samuel Kortum, Brent Neiman, and John Romalis. 2016. "Trade and the Global Recession." *American Economic Review* 106 (11): 3401–38.

Fajgelbaum, Pablo D., and Edouard Schaal. 2020. "Optimal Transport Networks in Spatial Equilibrium." *Econometrica* 88 (4): 1411–52.

Fally, Thibault, and James Sayre. 2018. "Commodity Trade Matters." NBER Working Paper 24965.

Fan, Jingting, Yi Lu, and Wenlan Luo. 2019. "Valuing Domestic Transport Infrastructure: A View from the Route Choice of Exporters." *Review of Economics and Statistics*. https://doi.org/10.1162/rest_a_01084.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer. 2022. "Penn World Table version 10.0." Groningen Growth and Development Centre [distributor] https://doi.org/10.15141/S5Q94M (accessed January 2023).

Feenstra, Robert C., Robert E. Lipsey, Haiyan Deng, Alyson C. Ma, and Hengyong Mo. 2005. "Data for: World Trade Flows: 1962–2000." NBER Working Paper 11040.

Feyrer, James. 2021. "Distance, Trade, and Income—The 1967 to 1975 Closing of the Suez Canal as a Natural Experiment." *Journal of Development Economics* 153: 102708.

Fransoo, Jan C., and Marc J. F. Wouters. 2000. "Measuring the Bullwhip Effect in the Supply Chain." *Supply Chain Management: An International Journal* 5 (2): 78–89.

Friedt, Felix L., and Wesley W. Wilson. 2020. "Trade, Transport Costs and Trade Imbalances: An Empirical Examination of International Markets and Backhauls." *Canadian Journal of Economics* 53 (2): 592–636.

Fuchs, Simon, and Woan Foong Wong. 2022. "Multimodal Transport Networks." Federal Reserve Bank of Atlanta Working Paper 2022-13.

Ganapati, Sharat. 2018. "The Modern Wholesaler: Global Sourcing, Domestic Distribution, and Scale Economies." Unpublished.

Ganapati, Sharat, and Woan Foong Wong. 2023. "Replication data for: How Far Goods Travel: Global

Transport and Supply Chains from 1965–2020." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E188301V1.

Ganapati, Sharat, Woan Foong Wong, and Oren Ziv. 2021. "Entrepôt: Hubs, Scale, and Trade Costs." NBER Working Paper 29015.

Gaulier, Guillaume, and Soledad Zignago. 2010. "Data for: BACI: International Trade Database at the Product-Level (the 1994-2007 Version)." CEPII Working Paper 2010-23.

Grant, Matthew, and Meredith Startz. 2022. "Cutting Out the Middleman: The Structure of Chains of Intermediation." NBER Working Paper 30109.

Harley, C. Knick. 1988. "Ocean Freight Rates and Productivity, 1740–1913: The Primacy of Mechanical Invention Reaffirmed." *Journal of Economic History* 48 (4): 851–76.

Harrigan, James. 2010. "Airplanes and Comparative Advantage." *Journal of International Economics* 82 (2): 181–94.

Hayakawa, Kazunobu, Jota Ishikawa, and Nori Tarui. 2020. "What Goes Around Comes Around: Export-Enhancing Effects of Import-Tariff Reductions." *Journal of International Economics* 126: 103362.

Head, Keith, and Thierry Mayer. 2013. "What Separates Us? Sources of Resistance to Globalization." *Canadian Journal of Economics* 46 (4): 1196–1231.

Heiland, Inga, Andreas Moxnes, Karen Helene Ulltveit-Moe, and Yuan Zi. 2019. "Trade from Space: Shipping Networks and the Global Implications of Local Shocks." CEPR Discussion Paper 14193.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2004. "Export versus FDI with Heterogeneous Firms." *American Economic Review* 94 (1): 300–316.

Hesse, Markus, and Jean-Paul Rodrigue. 2004. "The Transport Geography of Logistics and Freight Distribution." *Journal of Transport Geography* 12 (3): 171–84.

Hummels, David. 2007. "Transportation Costs and International Trade in the Second Era of Globalization." *Journal of Economic Perspectives* 21 (3): 131–54.

Hummels, David L., and Georg Schaur. 2013. "Time as a Trade Barrier." *American Economic Review* 103 (7): 2935–59.

Hummels, David, and Alexandre Skiba. 2004. "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture." *Journal of Political Economy* 112 (6): 1384–1402.

Hummels, David, Volodymyr Lugovskyy, and Alexandre Skiba. 2009. "The Trade Reducing Effects of Market Power in International Shipping." *Journal of Development Economics* 89 (1): 84–97.

Ignatenko, Anna. 2021. "Price Discrimination and Competition in International Transportation." Unpublished.

Irarrazabal, Alfonso, Andreas Moxnes, and Luca David Opromolla. 2015. "The Tip of the Iceberg: A Quantitative Framework for Estimating Trade Costs." *Review of Economics and Statistics* 97 (4): 777–92.

Jacks, David S., Christopher M. Meissner, and Dennis Novy. 2008. "Trade Costs, 1870–2000." *American Economic Review: Papers & Proceedings* 98 (2): 529–34.

Johnson, Robert C., and Guillermo Noguera. 2012. "Accounting for Intermediates: Production Sharing and Trade in Value Added." *Journal of International Economics* 86 (2): 224-36.

Jonkeren, Olaf, Erhan Demirel, Jos van Ommeren, and Piet Rietveld. 2011. "Endogenous Transport Prices and Trade Imbalances." *Journal of Economic Geography* 11 (3): 509–27.

Khanna, Gaurav, Nicolas Morales, and Nitya Pandalai-Nayar. 2022. "Supply Chain Resilience: Evidence from Indian Firms." NBER Working Paper 30689.

Lashkaripour, Ahmad. 2020. "Weight-Based Quality Specialization." *Journal of International Economics* 127: 103380.

Lee, Hau L., Venkata Padmanabhan, and Seungjin Whang. 1997. "The Bullwhip Effect in Supply Chains." *MIT Sloan Management Review* 38 (3): 93–102.

Levinson, Marc. 2016. *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger*. Princeton, NJ: Princeton University Press.

Limao, Nuno, and Anthony J. Venables. 2001. "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade." *World Bank Economic Review* 15 (3): 451–79.

Lugovskyy, Volodymyr, Alexandre Skiba, and David Terner. 2022. "Unintended Consequences of Environmental Regulation of Maritime Shipping: Carbon Leakage to Air Shipping." Unpublished.

Milne, Richard. 2013. "Nils Andersen, CEO, AP Møller-Maersk." *Financial Times*, October 6. https://www.ft.com/content/7c77131e-2c1f-11e3-8b20-00144feab7de.

Moore, Scott. 2010. "A Profile of Toyota's Production System." Harvard Business School Case Study

W90C18-PDF-ENG.

Mundaca, Gabriela, Jon Strand, and Ian R. Young. 2021. "Carbon Pricing of International Transport Fuels: Impacts on Carbon Emissions and Trade Activity." *Journal of Environmental Economics and Management* 110: 102517.

Nishiguchi, Toshihiro, and Alexandre Beaudet. 1998. "The Toyota Group and the Aisin Fire." *MIT Sloan Management Review* 40 (1): 49–60.

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

Pelton, Tom. 1994. "Hawthorne Works' Glory Now Just So Much Rubble." *Chicago Tribune*, April 18. https://www.chicagotribune.com/news/ct-xpm-1994-04-18-9404180047-story.html.

Petrova, Magdalena. 2018. "Inside Apple's iPhone: Where Parts and Materials Come From." *CNBC*, December 14. https://www.cnbc.com/2018/12/13/inside-apple-iphone-where-parts-and-materials-come-from.html.

Poole, Jennifer. 2016. "Business Travel as an Input to International Trade." Unpublished.

Proctor, Jon, Mike Machat, and Craig Kodera. 2010. *From Props to Jets: Commercial Aviation's Transition to the Jet Age 1952–1962*. North Branch, MN: Specialty Press.

Redding, Stephen J. 2022. "Trade and Geography." In *Handbook of International Economics: International Trade*, Vol. 5, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 147–217. Amsterdam: Elsevier.

Redding, Stephen J, and Matthew A Turner. 2015. "Transportation Costs and the Spatial Organization of Economic Activity." In *Handbook of Regional and Urban Economics*, Vol. 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 1339–98. Amsterdam, North-Holland.

Redding, Stephen J., Daniel M. Sturm, and Nikolaus Wolf. 2011. "History and Industry Location: Evidence from German Airports." *Review of Economics and Statistics* 93 (3): 814–31.

Rodrigue, Jean-Paul. 2020. *The Geography of Transport Systems*. Abingdon, UK: Routledge.

Rua, Gisela. 2014. "Diffusion of Containerization." FEDS Working Paper 2014–88.

Schlagheck, Dennis, and Catherine Lantz. 2014. *Hawthorne Works*. Charleston, SC: Arcadia Publishing.

Shapiro, Joseph S. 2016. "Trade Costs, CO2, and the Environment." *American Economic Journal: Economic Policy* 8 (4): 2020–54.

Stromberg, Joseph. 2015. "The MSC Oscar Just Became the World's Biggest Container Ship." *Vox*, January 8. https://www.vox.com/2015/1/8/7513317/container-ship-msc-oscar.

Swanson, Ana. 2021. "Crunch at Ports May Mean Crisis for American Farms." *New York Times*, November 14. https://www.nytimes.com/2021/11/14/business/economy/farm-exports-supply-chain-ports.html?login=email&auth=login-email.

Talley, Wayne K., and Sara Russell Riggs. 2018. "Trade Logistics and Seaborne Transportation." In *Handbook of International Trade and Transportation*, edited by Bruce A. Blonigen and Wesley W. Wilson, 409–21. Cheltenham, UK: Edward Elgar Publishing.

Tanaka, Kiyoyasu, and Kenmei Tsubota. 2016. "Directional Imbalance in Freight Rates: Evidence from Japanese Inter-prefectural Data." *Journal of Economic Geography* 17 (1): 217–32.

Timmer, Marcel P., Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J. de Vries. 2015. "Data for: An Illustrated User Guide to the World Input–Output Database: The Case of Global Automotive Production." *Review of International Economics* 23 (3): 575–605.

United Nations Statistics Division. 2022. "Gross Value Added by Kind of Economic Activity at Current Prices." New York: United Nations.

UNCTAD. 2021. *Review of Maritime Transport*. New York: United Nations Publications.

UNCTAD. 2022. "The Geography of Trade and Supply Chain Reconfiguration: Implications for Trade, Global Value Chains and Maritime Transport." Paper presented at the Thirteenth Session of the Trade and Development Commission, Geneva, November 21–25.

Van Leemput, Eva. 2021. "A Passage to India: Quantifying Internal and External Barriers to Trade." *Journal of International Economics* 131: 103473.

Weber, Austin. 2002. "The Hawthorne Works." *Assembly*, August 1. https://www.assemblymag.com/articles/88188-the-hawthorne-works.

Western Electric. 1938. "The Story of Western Electric: Service of Supply for the Bell System." https://worldradiohistory.com/Archive-Catalogs/Western-Electric/The-Story-of-Western-Electric-1938.pdf.

Woltjer, P. J., Reitze Gouma, and Marcel P. Timmer. 2021. "Data for: Long-Run World Input-Output

Database: Version 1.0 Sources and Methods." GGDC Research Memorandum 190.
Wong, Woan Foong. 2022. "The Round Trip Effect: Endogenous Transport Costs and International Trade." *American Economic Journal: Applied Economics* 14 (4): 127–66.
World Trade Organization. 2023. "World Trade Growth, 1950–2021." World Trade Organization. https://www.wto.org/english/res_e/statis_e/trade_evolution_e/evolution_trade_wto_e.htm (accessed on April 1, 2023)
Yilmazkuday, Demet, and Hakan Yilmazkuday. 2017. "The Role of Direct Flights in Trade Costs." *Review of World Economics* 153 (2): 249–70.

Journal of Economic Perspectives—Volume 37, Number 3—Summer 2023—Pages 31–58

The Changing Firm and Country Boundaries of US Manufacturers in Global Value Chains

Teresa C. Fort

There are two main concerns about US manufacturing and globalization: the loss of US jobs to foreign places and the loss of a US knowledge base connected to manufacturing. In this paper, I document the full range of US manufacturing firms' domestic and global operations, providing a broader context for these concerns.

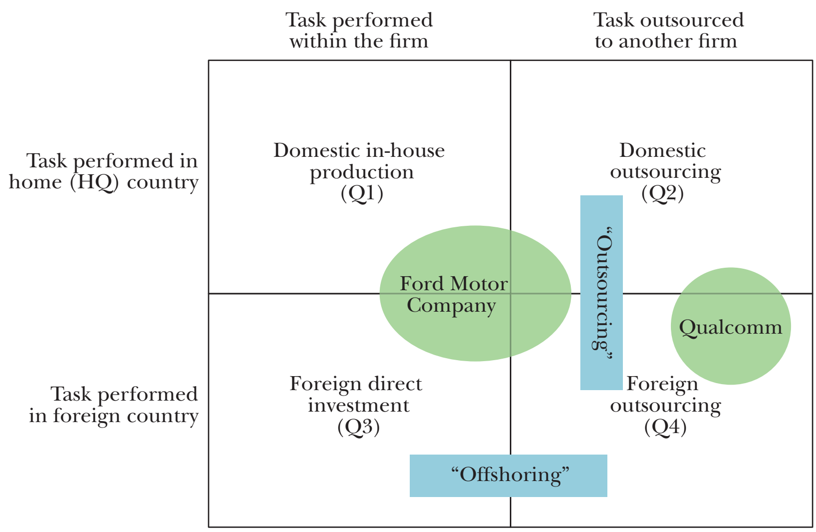
My perspective emphasizes that manufacturing involves three major stages: (1) product design and innovation; (2) a series of physical transformation activities, such as making inputs and assembling them; and (3) sales, marketing, and distribution. Most trade models implicitly (or even explicitly) include all three stages, but government statistics only classify physical transformation tasks as manufacturing. This mismatch from theory to data was not problematic when all three stages were performed inside a particular firm and country. However, dramatic improvements in information and communication technology have made it increasingly possible to fragment these stages across multiple countries and firms. This fragmentation has made measuring the complete production process for manufactured goods difficult (or even impossible) with traditional datasets.

Figure 1 illustrates a firm's choices to fragment production across countries and firms, building on a diagram introduced by Feenstra (2010). The horizontal axis captures the firm boundary decision for physical transformation tasks: the firm

■ *Teresa C. Fort is an Associate Professor of Business Administration, Tuck School of Business, Dartmouth College, Hanover, New Hampshire. She is a Faculty Research Fellow at the National Bureau of Economic Research, Cambridge, Massachusetts, and an Affiliate at the Centre for Economic Policy Research, London, United Kingdom. Her email address is teresa.fort@dartmouth.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.3.31.

Figure 1
Firm Integration and Location Decisions for Physical Transformation Tasks



Source: This figure builds on the framework presented in Feenstra (2010).
Note: The figure depicts the firm (horizontal axis) and country (vertical axis) boundary decisions made by firms involved in the broader manufacturing process for the stage 2, physical transformation tasks required to manufacture goods.

may "outsource" tasks to other firms or maintain integrated production. The vertical axis captures the firm's location choice for these tasks: the firm may "offshore" by locating production in one or more foreign countries, or produce domestically.

A US firm with manufacturing plants in the United States necessarily occupies quadrant 1, since it performs physical transformation tasks in-house. A firm that owns foreign manufacturing plants occupies quadrant 3. However, a single firm may occupy multiple quadrants. For example, consider Texas Instruments, a US semiconductor manufacturer that owns and operates multiple wafer fabs in the United States, along with eleven other production sites in Mexico, Europe, and Asia. The Ford Motor Company has 30 manufacturing plants, 20 of which are in foreign countries. Since these firms manufacture in-house in the United States and abroad, they span quadrants 1 and 3. Both Texas Instruments and Ford also work extensively with arm's-length partners, such that they also span quadrants 2 and 4, thus covering the entire matrix.[1]

In contrast to firms that perform physical transformation tasks in-house, a "factoryless goods producer" is a firm that contracts for all of its physical

[1] See https://www.ti.com/about-ti/company/ti-at-a-glance/manufacturing.html for TI plants, https://www.ti.com/about-ti/suppliers/supplier-portal.html?keyMatch=SUPPLIERS for TI supplier portal, https://corporate.ford.com/operations/locations/global-plants.html for Ford plants, and https://fsp.portal.covisint.com/web/portal for the Ford Supplier Portal.

transformation activities, and as such occupies only quadrants 2 and/or 4, depending on whether its suppliers are domestic, foreign, or both. For example, Apple is deeply involved in manufacturing physical goods, but does so by means of third-party contract manufacturing suppliers primarily in foreign countries. Similarly, Nike reports 640 manufacturing locations across 38 countries, all of which involve outsourced relationships with contract manufacturers. Qualcomm is one of many "fabless" semiconductor firms that design chips and rely on predominantly Korean and Taiwanese contract manufacturers for their production; indeed, Bayard, Byrne, and Smith (2015) attribute 25 percent of global semiconductor sales in 2012 to such factoryless goods producers.[2]

It is worth clarifying that fragmenting production does not just refer to purchasing inputs. For example, a Belgian candymaker's imports of chocolate are generally not considered fragmented production in studies on outsourcing and offshoring. Instead, fragmented production entails a splitting apart of the production process by a firm that used to produce, or could reasonably have produced, the fragmented part.

Firms with no domestic manufacturing plants have no activity in quadrant 1. They tend to be missing from research on production fragmentation and offshoring because there is no clear way in standard datasets to identify their direct involvement in manufacturing. In this paper, I begin by describing the limitations of standard datasets in identifying such firms. I then exploit two novel US data sources to identify two organizational forms missing from many analyses on global value chains: US firms that perform physical transformation tasks within the firm boundary using exclusively foreign manufacturing plants, and factoryless goods producers that outsource all their physical transformation tasks to arm's-length contract manufacturers.

Contrary to the fear that US multinationals have offshored most of their jobs, I find that the vast majority of US firms that own foreign manufacturing plants in 2007 also maintain domestic production; moreover, manufacturing comprises their primary domestic activity. Contrary to the fear that participation in global value chains entails a loss of technological skills, I find that firms with global in-house manufacturing plants and factoryless goods producers both employ relatively high shares of US "knowledge" workers. Indeed, multinational enterprises that manufacture goods are disproportionate contributors to R&D and patenting, and factoryless goods producers are far more likely to design goods than other firms in their sector, and have also been linked to greater R&D, patenting, and trademarks (Kamal 2023).

A complete picture of US firms' involvement in global value chains is necessary to understand the effects of globalization. For example Berman, Bound, and Griliches (1994) ruled out trade as an explanation for the shift towards nonproduction workers in US manufacturing plants during the 1980s and 1990s because those changes occurred in some of the same industries with surging imports, notably

[2] For Apple, see https://www.wsj.com/articles/BL-DGB-25630; for Nike, see https://manufacturingmap.nikeinc.com/#; and for Qualcomm, see https://www.qualcomm.com/company/corporate-responsibility/acting-responsibly/sustainable-product-design/supply-chain.

computer and electronics. They reasoned that a trade explanation must entail reallocation across industries consistent with US comparative advantage, whereas within-industry adjustments dominated in the data. We now know, however, that computer and electronic manufacturing is one of the first industries in which factoryless good production arose, as some firms focused on innovation and shifted physical production to foreign suppliers. It is notable that computer and electronics also accounts for the greatest growth in breakthrough patents over the last two decades (Kelly et al. 2021) and the majority of real value added growth in US manufacturing from 1992 to 2011, even as imports of computers and electronics surged (Fort et al. 2017).

I conclude with a discussion on how trade statistics and theory need to expand to capture the realities of goods production across firm and country boundaries. These activities affect our understanding of trade and foreign direct investment, as well as aggregate measures of domestic value added and GDP. The potential implications are far-reaching: increased specialization within the production of a particular industry or good provides additional gains from trade (Jones and Kierzkowski 2001). When such specialization entails reallocation into early production stages, like design and innovation, offshoring can even lead to dynamic gains, as the returns to innovation rise, inducing growth in R&D and ideas that beget more ideas (Grossman and Helpman 1991; Rodríguez-Clare 2010).

Measuring the Range of Manufacturers' Organizational Forms

The traditional data on manufacturing firms collected by US statistical agencies make it difficult to capture firms that are involved in the manufacturing process by producing goods exclusively outside the United States, exclusively outside their firm boundaries by means of contracts, or both. I review the existing data on US establishments and firms, on international trade by firms, and on multinational firms, explaining the benefits and limitations of each data source.

Standard Measures of US Establishments and Firms

The Census Bureau defines an "establishment" as a physical location at which employment and payroll records are kept. A firm can thus have multiple establishments—and these establishments need not be classified in the same industry.

The US Census Bureau constructs the Longitudinal Business Database, which is a comprehensive, establishment-level dataset of all private, nonfarm employer establishments from 1976 to 2019. The dataset provides employment, payroll, location, and a unique industry code for the primary activity of every establishment (for details, see Jarmin and Miranda 2002; Chow et al. 2021). All employees of an establishment are assigned to its industry. The data also identify the firm to which each establishment belongs, making it possible to measure the full range of a firm's activities across industries and sectors.

The Longitudinal Business Database can be merged to the Economic Censuses, which are collected in years that end in 2 or 7. These censuses contain detailed information on establishment sales, input use, and other sector-specific metrics, such as technology.

An establishment's industry is the primary means that government agencies and researchers use to identify manufacturing activity. US statistical agencies use the North American Industry Classification System, commonly referred to as NAICS (and described at https://www.census.gov/naics) to classify establishments. The guiding principle of NAICS is to assign an industry code to an establishment based on the main activities performed by its employees. By contrast, the earlier Standard Industrial Classification System (SIC) classified establishments that provided support services for other establishments of their firm to those establishments' industry. For example, an R&D lab is always in Services under NAICS, but would have been classified in manufacturing under SIC if its R&D supported the firm's manufacturing plants. US Census data transitioned from NAICS to SIC between 1997 to 2002, a period that coincides with China's entry to the World Trade Organization, making this issue particularly relevant for research on globalization. The current Longitudinal Business Database now includes the most recent vintage of NAICS codes for every establishment over the entire period using the methods developed in Fort and Klimek (2018).

Factoryless goods producers are hard to identify with these data. To be classified in manufacturing, an establishment must perform "mechanical, physical, or chemical transformation of materials or components into new products." As a result, an establishment that contracts for manufacturing services will generally be classified in nonmanufacturing sectors, such as wholesale trade.[3] Similarly, establishments that perform support activities for manufacturing, such as an R&D lab or an engineering services establishment, are classified in services. Given the ongoing fragmentation of design and production, we need new ways to link the contributions of US service workers to the value of the manufactured goods they design and market.

Merchandise Trade Data by Firm

Starting in 1992, the Census firm data can be merged with firm-level data from US Customs that record the universe of trade transactions above $2,500 of merchandise goods that enter or leave the United States. This Longitudinal Firm Trade Transactions Database (LFTTD) provides detailed information on the products shipped, as well as the source (for imports) or destination (for exports). These data were first linked by Bernard, Jensen, and Schott (2009); Kamal and Ouyang (2020) provide details on the latest linking efforts.

One unique feature of these US trade data is that they contain an identifier for transactions between related parties. Export transactions in which one of the parties has at least 10 percent ownership of the other party, or imports between parties with a 5 percent ownership threshold are flagged as "related-party" transactions.[4]

[3] For an explanation from the Census Bureau, see https://www.census.gov/naics/?input=31&year=2022&details=31. An exception is "jobbers" in certain apparel manufacturing industries. These establishments perform the "entrepreneurial functions involved in apparel manufacturing," but contract for the transformation activities from other firms.

[4] Related-party imports may include other relationships. For details, see Kamal and Ouyang (2020).

Thus, arm's-length trade can be distinguished from flows between related parties. However, for the present purpose of studying whether US firms have integrated manufacturing plants in foreign countries, these data have well-known shortcomings: they have no information on activities of affiliates of multinational enterprises; they do not distinguish US multinational enterprises from foreign-owned firms; and they are based on very low ownership thresholds.

Foreign Direct Investment Data by Firm

The US Bureau of Economic Analysis carries out the Annual Survey of US Direct Investment Abroad, known as BE-11, which provides information on all US-based firms' outward foreign affiliate employment, local sales, sales back to the United States (and whether these are intrafirm), and sales to third markets, by the affiliate country and industry. This survey thus captures outward foreign direct investment information. The Bureau of Economic Analysis also carries out the Benchmark Survey of Foreign Direct Investment in the United States, known as BE-12, which is conducted every five years and provides inward foreign direct investment. This survey makes it possible to identify foreign-owned firms operating in the United States.

For 2007, the year of my analysis, these two surveys provide the most detailed and comprehensive information available about multinational firms operating in the United States. In contrast to the Census Bureau's related-party trade data, these data include share-of-ownership information, as well as foreign affiliates' industries and their local, US, and third-market sales.

Despite their advantages, these data alone are not sufficient to study all of US firms' manufacturing activities. First, these data only include multinationals, so there is no information on domestic manufacturers. Second, they are reported at the firm level, and therefore do not contain the establishment-level information necessary to analyze the full range of firms' domestic establishments. Finally, the data lack country- and product-level information on the universe of firms' imports and exports.

In the next section, I combine the 2007 US Census Bureau and Bureau of Economic Analysis data described here to identify all US firms with integrated manufacturing plants anywhere in the world. Such an analysis ensures coverage of all firms with any activity in quadrants 1 or 3 of Figure 1. In the subsequent section, I exploit detailed questions from the 2017 Economic Census of Wholesale Trade to identify firms that are involved in the broader manufacturing process by contracting for production from arm's-length suppliers, thus capturing firms specialized in quadrants 2 and/or 4.

Country Boundaries of Integrated US Manufacturers

In this section, I focus on US firms that are directly involved in manufacturing because they have majority-ownership shares in manufacturing plants in the United States, in foreign countries, or both.

As such, these firms necessarily occupy quadrants 1 or 3 of Figure 1. In the next section, I turn to factoryless goods producers such as Apple and Nike.

Novel Data on All In-House Manufacturing by US Firms

I use new data merged and analyzed by Kamal, McCloskey, and Ouyang (2022) and Antràs et al. (2023) to provide a complete picture of US firms that perform physical transformation tasks in-house anywhere in the world. A key contribution of my analysis is to include firms with no domestic manufacturing plants, which are missing from studies using traditional datasets. I measure firms' employment, sales, and trade activity across sectors by linking the 2007 Longitudinal Business Database, Economic Censuses, and the Longitudinal Firm Trade Transactions Database (excluding trade of minerals, fuels, and oil [HS 27]). I identify multinational enterprises as all US firms with majority-owned foreign affiliates using the "outward" foreign direct investment survey. I use the "inward" survey to remove all US establishments that are majority-owned by a foreign firm.

In 2007, there are 243,700 US firms that own manufacturing plants somewhere in the world (about 5.6 percent of all firms), which account for 88 percent of total US manufacturing employment (foreign multinational enterprises employ the remainder), 20 percent of total employment, and 29 percent of total US sales. These firms mediate 42 percent of US imports and 58 percent of exports, which highlights the disproportionate involvement of goods-producing firms in international trade. Appendix Table A.1 provides a more detailed decomposition of these statistics.

Domestic versus Offshored Integrated Manufacturing

A common perception is that US multinational enterprises have relocated the bulk of their manufacturing plants offshore. To evaluate this claim, I use the new data to categorize all US firms with majority-owned manufacturing plants anywhere in the world into four categories: (1) domestic manufacturing firms without any majority-owned foreign affiliates; (2) US multinational enterprises that have only US manufacturing plants (their foreign affiliates are outside manufacturing); (3) US multinational enterprises that have both US and foreign manufacturing plants; and (4) US multinational enterprises that have only foreign manufacturing plants. (All these firm types may also outsource some tasks from domestic or foreign suppliers.)

The first row of Table 1 presents the number of US firms that manufactured in-house in 2007 across these four categories. Of the 243,700 US manufacturing firms, only 1,700 have majority-owned foreign establishments (columns 2 to 4). Among these multinationals, 1,200 firms own US and foreign manufacturing plants versus 350 firms with just domestic plants and only 150 firms with exclusively foreign in-house manufacturing. Firms with both domestic and foreign manufacturing plants are thus the most prevalent type of US multinational manufacturing enterprise.

Panel A of Table 1 presents total sales for these firms. The first row contains global sales, which are the sum of firms' US and foreign-establishment sales, each of which is presented separately in the next two rows. I include firms' total sales here, regardless of whether they are booked by manufacturing or nonmanufacturing establishments.

Table 1
Sales, Employment, and Trade Flows for All US Firms that Manufacture In-house in 2007

Firm Type: *Majority-Owned Manufacturing Plants In:*	*Domestic* *US Only* (1)	*Multinational Enterprises* *US Only* (2)	*Multinational Enterprises* *US & Foreign* (3)	*Multinational Enterprises* *Foreign Only* (4)
Firms	242,000	350	1,200	150
Panel A. Sales ($billions)				
Global Sales	2,629	1,695	6,710	345
Sales by US Establishments	2,629	1,446	3,853	173
Sales by Foreign Establishments	—	249	2,857	172
Panel B. Employment (thousands)				
Global Employment	11,059	5,338	11,883	732
Employment in US Establishments	11,059	4,349	6,556	361
Employment in Foreign Establishments	—	989	5,327	371
Panel C. US Trade Flows ($billions)				
Imports	126	39	410	12
Arm's-Length	89	33	160	6
Related-Party	37	7	250	6
Exports	123	22	437	3
Arm's-Length	103	16	253	2
Related-Party	19	5	184	1

Source: 2007 Longitudinal Business Database, Economic Censuses, Longitudinal Firm Trade Transactions Database, BEA inward and outward surveys.
Note: Table presents total number of firms and their global sales, global employment, and US merchandise good trade flows in 2007 by firm type and manufacturing plant locations. Sample is all US firms with one or more majority-owned manufacturing plants anywhere in the world in 2007.

The sales data deliver two stark messages. First, US firms with both domestic and foreign manufacturing plants dominate both global and US sales, with global sales of $6.7 trillion—more than the other three categories combined—despite the fact that they are only 1,200 out of the 243,700 firms in the sample. Second, US manufacturers that only produce in-house in foreign plants account for a mere 3 percent US manufacturers' global sales.

The dominance of firms with both US and foreign in-house production is reinforced by firms' employment differences. Panel B shows that transnational manufacturers—those that perform in-house physical transformation activities in the United States and abroad—employ more workers than all other firm types, with just over half of these workers employed at their US plants. Firms that manufacture exclusively in foreign plants employ less than one million workers worldwide and account for just 2.5 percent of all US manufacturing firms' global employment. In short, the notion that US firms moved almost all of their integrated manufacturing plants overseas in response to China's accession to the World Trade Organization in 2001, and then used those plants to serve their US customers, is not supported by the 2007 data. Instead, when a firm integrates physical transformation tasks, it also maintains domestic production.

I also assess the importance of manufacturing for these firm types. Table 2 shows that domestic manufacturers are the most specialized in physical transformation

Table 2
US Manufacturers' Sales and Employment Shares by Sector and Establishment Locations

Firm Type:	*Domestic*	*Multinational Enterprises*	*Multinational Enterprises*	*Multinational Enterprises*
Majority-Owned Manufacturing Plants In:	*US Only*	*US Only*	*US & Foreign*	*Foreign Only*
	(1)	*(2)*	*(3)*	*(4)*
Global Manufacturing Sales/Global Sales	0.69	0.10	0.57	0.07
US Establishments	0.69	0.12	0.54	—
Foreign Establishments	—	—	0.60	0.14
Global Manufacturing Employment/Global Employment	0.69	0.06	0.66	0.11
US Establishments	0.69	0.08	0.55	—
Foreign Establishments	—	—	0.79	0.21
US Professional & Management Employment/US Employment	0.03	0.10	0.19	0.15

Source: 2007 Longitudinal Business Database, Economic Censuses, Longitudinal Firm Trade Transactions Database, BEA inward and outward surveys.
Note: Table presents shares of firm sales and employment in manufacturing establishments for all establishments, and by US or foreign establishments. Bottom row presents US establishment employment in Professional, Scientific, and Technical Services (NAICS 54) and Management (NAICS 55) over total US employment. Sample is all US firms with one or more majority-owned manufacturing plants anywhere in the world.

tasks, with 69 percent of their sales and employment in manufacturing plants. Firms that manufacture in the United States and abroad have the next highest share, with 57 percent of their global sales and 66 percent of employment in manufacturing plants. By contrast, US multinationals that manufacture exclusively in the United States or exclusively abroad have manufacturing sales and employment shares that range from just 8 to 21 percent. Among US multinationals that manufacture goods, physical transformation tasks are thus only a significant activity for those with manufacturing plants both at home and abroad.

Table 2 also shows that the majority of US sales and employment for these transnational manufacturers is in manufacturing plants; specifically, 55 percent of their US employees work in manufacturing plants, compared to 79 percent of their foreign workers. Their foreign workforce is thus geared more towards production work, but these firms still maintain physical transformation tasks as their primary US activity. Finally, I use the data from Tables 1 and 2 to calculate that transnational manufacturers' US plants account for 55 percent of their total manufacturing sales and 46 percent of their global manufacturing employment.

To summarize, even seven years after China's accession to the World Trade Organization, US multinationals that manufacture in-house tend to do so in both the United States and foreign countries, and their US manufacturing plants comprise the majority of their domestic activities. These firms' global manufacturing activities are roughly split across their US and foreign plants, with just over half of their total manufacturing plant sales originating from US establishments, and just under half of their manufacturing plant workers located in the United States. These patterns highlight a potential interdependence between the organizational and national boundary decisions of US manufacturers: "in the firm" also entails a substantial share "in the home country."

Relationship between Domestic and Foreign Production

Transnational manufacturers also dominate trade flows. Panel C of Table 1 presents exports and imports for the same four categories of firms. Transnational manufacturers import $410 billion in goods and export $437 billion, which is almost four times the amount of either trade flow for all the other firm categories combined. Their trade dominance is due not only to their size, but also to their disproportionate trade intensity. The ratio of total exports to sales for these firms is 0.10, compared to 0.05 for nonmultinational enterprises and 0.01 for US multinational enterprises that only manufacture in the United States.

Recall that the trade data distinguish between intra- and across-firm transactions. Unlike the ownership threshold of at least 50 percent I used to classify the status of multinational enterprises, a "related party" in the trade data denotes exports to partners with a 10 percent ownership threshold or imports from parties with a 5 percent ownership threshold. As such, it is possible for domestic firms to engage in related-party trade, which they do to some extent. Panel C of Table 1 decomposes imports and exports along these lines and shows that the majority (0.58) of transnational manufacturers' exports go to arm's-length partners. These arm's-length shipments suggest that multinational enterprises' US manufacturing plants also serve foreign customers.

Further insight into the motives for multinationals' US exports can be gained by studying the countries to which they sell. Exploiting the novel country-level trade dimension of these merged data, Antràs et al. (2023) show that US multinationals are much more likely to export not only to the countries in which they have affiliates, but also to countries that are proximate to their affiliates or that share a free trade agreement with them. Those authors use a framework in which firms must incur a fixed cost to sell their goods in a particular country; for example, to learn about a country's legal institutions, its demand (Foster, Haltiwanger, and Syverson 2008), or to locate specific customers (Bernard et al. 2022). They show that when this fixed cost is shared by all of the multinational firm's manufacturing plants, a firm's US plants will be more likely to export to markets that are proximate to its affiliates. This tilting arises because countries that are proximate to a foreign affiliate enjoy lower bilateral trade costs with the affiliate, thus increasing the marginal benefit of activating the market.

The same intuition applies to a firm's decision to source inputs. If the country-specific fixed cost to find suppliers and source inputs from a particular country is shared across all of the firm's plants, then firms with domestic and foreign production plants will source from more countries and use more imported inputs. The data indicate that transnational manufacturers are also the most import-intensive, with a ratio of imports to sales of 0.11, which is again more than double the ratio for domestic firms. As for exports, Antràs et al. (2023) show that the number of countries from which US manufacturers import is increasing in the number of foreign countries in which they manufacture, and that multinationals are more likely to import from countries that are proximate to their foreign production plants.

Of course a transnational manufacturing firm's US imports need not consist solely of inputs. Indeed, Table 1 shows that 61 percent of transnational manufacturers' imports are from related parties, which could be inputs or final goods produced by affiliates. For example, Ford produces sports utility vehicles in the United States, but imports its Fiesta models from Mexico. I use information from the Census of Manufacturers "product and material trailer files" to identify goods that the firms' US establishments produce and inputs that they purchase. I compare these goods and inputs to the products firms import (using the Customs data) and find that a significant portion of multinationals' imports consists of the same goods they manufacture in the United States, while another large share appears to be both produced goods and inputs.[5] This overlap, however, may reflect the fact that US trade, input, and production data are all collected using different classification systems and concording across them requires aggregating the data such that we can no longer distinguish an input from an output.

Related evidence from Danish data, however, suggests that a large portion of the apparent overlap at the coarse industry level represents trade of the same goods produced by the firm at home. Using a novel offshoring survey along with detailed production and import data that are collected using the same classification system, Bernard et al. (2023) find that Danish firms grow their imports of the same products they manufacture at home when they relocate production to low-wage countries. (Those authors also show that these imports of domestically produced goods appear to be inputs when aggregating the data.) The Danish firms continue domestic production of the imported goods, but the domestic varieties have higher unit values that grow after importing begins, consistent with firms producing lower quality or less technologically advanced varieties in lower-wage countries.

An interesting venue for future work is to assess whether US manufacturers similarly use their global production plants to produce vertically differentiated products in different countries. This type of vertical differentiation contrasts with the standard "proximity-concentration" tradeoff at the heart of many models about foreign direct investment, in which a US firm chooses to serve a particular market either through exports or a plant in the foreign market; it may also explain recent evidence that a US multinational's affiliates in one foreign country do not seem to compete with its affiliates in other countries (Garetto, Oldenski, and Ramondo 2019). Perhaps most exciting is the possibility that this "vertical offshoring" may foster innovation up the quality ladder (for example, as shown for Japan by Braguinsky et al. 2021), thus providing a new way in which globalization allows firms to push out the knowledge frontier. Indeed, Bernard et al. (2023) show that Danish firms with new production-cost savings opportunities in Eastern

[5] Ramondo, Rappoport, and Ruhl (2016) use the outward multinational data and find that intrafirm shipments from affiliates to their US parent are rare and do not seem to comprise inputs, though their data lack the full range of a US firm's activities and rely on input-output tables to identify inputs.

Europe reallocate their domestic workforce into R&D and technology occupations. US firms with an expertise in manufacturing goods may also leverage their domestic design capabilities by manufacturing similar goods across multiple countries.

Leveraging Knowledge Workers around the World

To assess the extent to which transnational manufacturers' domestic employment is in "knowledge-related" activities such as design and marketing, I calculate firms' total employment in establishments classified in Professional, Scientific, and Technical Services (NAICS 54) and Management (NAICS 55). These two sectors capture workers in knowledge-intensive activities, such as R&D, as well as marketing. The last row of Table 2 depicts US manufacturing firms' share of workers in these sectors. Consistent with prior evidence on the importance of multinational enterprises in innovation, the employment shares of US multinationals in these sectors are substantially higher than domestic firm shares. While domestic firms have only 3 percent of their total employment in Professional Services or Management establishments, multinationals' shares range from 10 to 19 percent. US multinationals with both domestic and foreign manufacturing plants have the highest share across all firm types: 19 percent of their employment is in these "knowledge" establishments, consistent with them performing pre- and postproduction tasks in the United States and leveraging their expertise to manufacture across multiple countries.

US firms with domestic and foreign manufacturing plants maintain manufacturing as their primary domestic activity. In contrast to canonical models of horizontal foreign direct investment, in which firms serve foreign markets through exports or foreign affiliates, they use their US plants to serve markets that are close to their foreign plants and ship goods from their foreign plants back to the United States. These patterns, along with evidence from Danish firms, suggest that US firms with integrated global manufacturing have a core competence in manufacturing particular goods that they leverage around the world with support from their US "knowledge" workers. By contrast, US manufacturers with exclusively foreign manufacturing plants are small in number, employment, sales, and trade flows. "In the firm" goes hand-in-hand with a significant portion also "in the country."

New Facts and Patterns on Factoryless Goods Producers

Factoryless goods producers differ from in-house manufacturers because they outsource all physical transformation activities to other firms. Although this type of firm includes examples as prominent as Apple, Nike, and Qualcomm, they are hard—or even impossible—to identify using standard datasets. Because these firms' establishments do not perform physical transformation activities themselves, they are classified in sectors such as Retail, Wholesale, and Professional Services, and generally cannot be distinguished from other establishments in those sectors that have no involvement with the broader manufacturing process.

Statistical agencies across the world understand the current data limitations and have undertaken significant efforts to measure contract manufacturing and factoryless goods production. The US Census Bureau began asking establishments in the 2002 Census of Wholesale Trade about their involvement in product design and use of contract manufacturing, and continued this practice in the 2007, 2012, and 2017 surveys. In some years, the Census also asked about purchases of contract manufacturing services in some of the Census of Services and in its annual Company Organization Survey sent to large, multi-unit firms. Unfortunately, the questions and samples are sufficiently different across years to make time series analyses infeasible.

In 2010, the US Office of Management's Economic Classification Policy Committee recommended classifying a factoryless goods producer as a firm that "outsources all transformation steps that traditionally have been considered manufacturing, but undertakes all of the entrepreneurial steps and arranges for all required capital, labor, and material inputs required to make a good" (Office of Management and Budget 2011). Moreover, the committee recommended reclassifying establishments that performed those related tasks into manufacturing for the 2012 Economic Census (Doherty 2015) to facilitate collection of additional information about use of inputs and sales by product, which are already part of the Census of Manufactures survey questions. However, this proposal was met with strong opposition from the US manufacturing lobby, and the reclassification effort was abandoned.[6] The Census Bureau has continued some of its data collection efforts for identifying factoryless goods producers, which I exploit in this paper.

Novel Data

I define a factoryless goods producer as a firm with no US manufacturing plants, but that is nevertheless involved in producing goods by contracting for manufacturing from other firms. I obtain data on firms' use of contract manufacturing using the 2017 Census of Wholesale Trade, which is sent to all establishments in the wholesale trade sector (NAICS 42). Wholesale establishments are traditionally intermediaries: they sell goods to other firms rather than to consumers, and they do not manufacture or transform the goods they sell. Wholesale establishments are primarily classified into two general categories: merchant wholesalers that buy and sell goods for other firms and manufacturing sales branches that sell merchandise manufactured by other establishments in their firm. I focus only on firms without in-house manufacturing plants, which cover 89 percent of firms in the 2017 published totals for the Wholesale Trade Sector and 68 percent of their employment.[7]

[6] For example, the director of industry research and technology at the Precision Machined Products Association stated, "We think it would be bad for policy makers to say, 'Look at these numbers, we have great manufacturing.'" See https://www.wsj.com/articles/SB10001424052702303546204579439170777269630.

[7] Although wholesale establishments are often warehouses, they differ from establishments classified as "warehouses" (NAICS 493) because wholesale establishments are responsible for the sale for their goods, whereas warehouses simply store merchandise, perhaps providing logistics and distribution support. See Appendix Section B.1 for additional details. The exact questions from the Census of Wholesale Trade that I use are presented in Appendix Figure C.1.

I exploit several questions from the "Special Inquiries" section in the 2017 Census of Wholesale Trade that ask whether the establishment had any manufacturing done on its behalf by other companies inside the United States and/or by other companies outside the United States; and whether the establishment determined the design or specifications of the products that were manufactured on its behalf. These questions thus capture purchases of contract manufacturing services by wholesale establishments, which are precisely the services purchased by firms like Apple and Nike that design their products and coordinate the production process, but locate physical transformation activities outside their firm boundary.

To analyze the activities of these firms across sectors and over time, I merge the 2017 Census of Wholesale Trade data to a panel of establishment-level employment and sales by sector from 1992 to 2017 using the Longitudinal Business Database and other Economic Censuses. I aggregate these data to the firm level and augment them with yearly firm-level imports and exports from the Longitudinal Firm Trade Transactions Database (recall that 1992 is the first year for which the Customs Trade data are available). I limit the sample to firms without any manufacturing establishments in 2017 and with at least one wholesale establishment that responded to one or more of the contract manufacturing questions in the Census of Wholesale Trade in 2017. Although I am missing these firms' foreign operations, the results in the prior section provide reassuring evidence that firms without US manufacturing plants tend not to have foreign manufacturing plants.

This sample of factoryless goods producers covers approximately half of all firms (and employment at firms) with one or more wholesale establishments and no manufacturing plants in 2017 (for details, see Appendix Table B.1). Thus, the sample is sufficiently large to perform a meaningful comparison of factoryless goods producers and their characteristics, but cannot be used to assess the aggregate importance of this organizational form. This limitation arises not only because approximately half of the wholesale sector is outside the sample, but also because factoryless goods producers may exist in other sectors for which the Census Bureau has not collected comparable data.

Characteristics of Factoryless Goods Producers

I first compare factoryless goods producers to the traditional merchant wholesalers in my sample, which do not contract for production from other firms. Again, this sample excludes all firms with any US manufacturing plants in 2017, and the earlier data on multinationals suggests that firms with exclusively foreign manufacturing plants are rare.

Table 3 presents weighted averages of firm characteristics for factoryless goods producers (27 percent of the sample) and for the rest. Factoryless goods producers are smaller on average than traditional wholesalers, with a weighted average of 26 workers per firm compared to 41 workers at other wholesalers. The average wage of factoryless goods producers is over 35 percent higher than the comparison group, and their sales per worker is over 10 percent larger. At the factoryless goods producers, 75 percent of the workforce is in wholesale establishments and 11 percent is in retail stores; by contrast, traditional merchant wholesalers have

Table 3
Characteristics of 2017 Wholesale Firms by Factoryless Goods Production Status

				Share of employment in	
	Average employment	*Pay/ Employment*	*Sales/ Employment*	*Wholesale*	*Retail*
Factoryless Goods Producers	26	76	773	0.75	0.11
Merchant Wholesalers	41	56	696	0.46	0.25
	Exports/Sales	*RP Exports/ Exports*	*Imports/Sales*	*RP Imports/ Imports*	*China Imports/ Imports*
Factoryless Goods Producers	0.05	0.31	0.25	0.49	0.36
Merchant Wholesalers	0.03	0.19	0.05	0.31	0.27

Source: 2017 Longitudinal Business Database, Longitudinal Firm Trade Transactions Database, Economic Census, Special Inquiry data from the Census of Wholesale Trade.
Note: Table presents weighted averages for factoryless goods producers and merchant wholesalers. Factoryless goods producers are firms that contract for manufacturing services from other firms. Sales in $1,000s. RP stands for "related party." Sample consists of all firms with one or more wholesale establishments that responded to the 2017 special inquires questions on contract manufacturing and with no US manufacturing plants. There are 37,300 factoryless goods producers and 99,500 merchant wholesalers in the sample.

46 percent of their employment in wholesale establishments and one-quarter in retail outlets.

Factoryless goods producers are also more trade-intensive than traditional wholesalers in the sample. The bottom panel of Table 3 presents these firms' exports-to-sales and imports-to-sales ratios, and their shares of related-party trade. Most notably, the imports-to-sales ratio is 0.25 for factoryless goods producers, compared to just 0.05 for the comparison group. This ratio of imports-to-sales for factoryless goods producers is also more than double the ratio of 0.11 at firms with transnational manufacturing plants. Factoryless goods producers also have higher import shares from China than traditional wholesalers: over one-third of their imports are Chinese.

Evolution of Factoryless Goods Producers

Prior work finds that factoryless goods producers tend to be younger (Bernard and Fort 2015; 2017), suggesting that the prevalence of this organization form may be growing. To investigate this possibility, I classify the 2017 firms in Table 3 into cohorts based on the first Economic Census year in which they enter the data, starting in 1992. Factoryless goods producers become more prevalent and have higher shares of employment in the later cohorts. Table 4 shows that 10 percent of the 2017 employment in factoryless goods producers is accounted for by firms that were born between 2012 and 2017, versus just 5 percent for merchant wholesalers. Traditional wholesalers are more likely to have entered prior to 2002: 75 percent of their 2017 employment is in firms that were alive by 1997, compared to only 58 percent for factoryless goods producers. Table 4 also shows that, at least since 2007, factoryless goods producers are similarly sized to traditional wholesalers within their same cohort. The average size of both types of firms born between

Table 4
Firms and Employment in 2017 by Factoryless Goods Production Status and Entry Cohort

	Factoryless Goods Producers			Merchant Wholesalers		
Entry Cohort	Firms	Share of Employment	Average Employment	Firms	Share of Employment	Average Employment
1992	7,700	0.50	62	27,000	0.65	99
1997	3,700	0.08	21	10,000	0.10	39
2002	4,500	0.10	21	12,000	0.08	27
2007	6,000	0.12	19	14,500	0.07	19
2012	6,500	0.11	17	15,500	0.06	15
2017	8,900	0.10	10	20,500	0.05	10
Totals	37,300	1.00	26	99,500	1.00	41

Source: 2017 Longitudinal Business Database, Longitudinal Firm Trade Transactions Database, Economic Census, Special Inquiry data from the Census of Wholesale Trade.
Note: Table presents the number of firms, share of employment, and average employment in 2017 based on firms' 2017 factoryless goods production status and the first Economic Census year in which they are alive. Firms listed in 1992 enter before or in 1992. Firms in the 1997 entry cohort enter between 1992 and 1997, and so on.

2012 and 2017 is just ten workers. The smaller size of factoryless goods producers in Table 3 is thus at least partly due to the fact that these firms are younger.

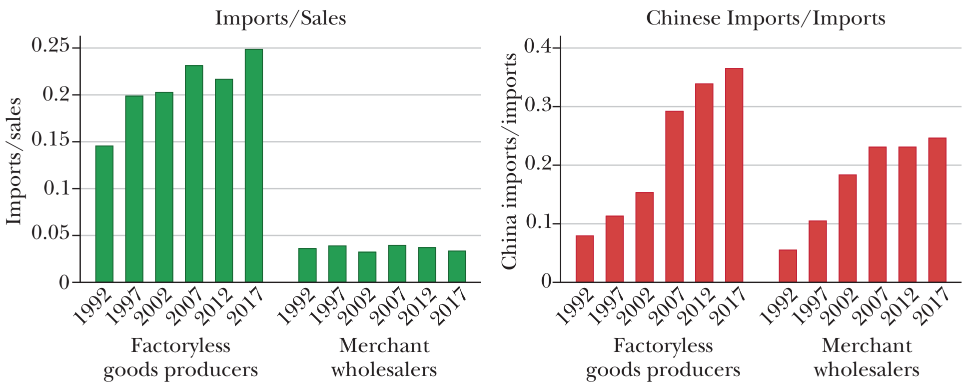
To assess whether factoryless goods producers and merchant wholesalers evolve differently, I trace the 2017 firms in my sample back in time, focusing only on those firms that also existed in 1992 (the firms in the first row of Table 4). While factoryless goods producers among these early entrants may still be younger, limiting the analysis to the subset of firms aged 25 and over reduces the selection effects due to differences in firm age.

Figure 2 reveals stark differences between factoryless goods producers and traditional wholesalers' import intensity that grows over time. The 2017 factoryless good producers that were alive in 1992 start with a high import intensity (0.15) in 1992 that grows 10 percentage points to reach 0.25 by 2017. By contrast, the 2017 merchant wholesalers also alive in 1992 maintain an imports-to-sales ratio below 0.05 throughout the period. The right panel of Figure 2 shows that factoryless goods producers are also more specialized in trade from China. Although the two types of firms have similar shares of imports from China in early years, the 2017 factoryless goods producers experience a much sharper increase following China's accession to the World Trade Organization in 2001. Factoryless goods producers are thus more outwardly oriented, with a larger share of their imports from China, one of the top low-wage manufacturing locations in the world.

I also use the data on 2017 firms that were alive by 1992 to analyze how firms' employment across sectors has evolved over time. Figure 3 presents the distribution of firms' employment across Wholesale (NAICS 42), Manufacturing (NAICS 31-33), Retail (NAICS 44–45), and Professional, Scientific, and Technical Services and Management (NAICS 54–55) sectors. Recall that, by definition, firms in the sample have no manufacturing employment in 2017.

Figure 2

Import Activity for 2017 Firms by Factoryless Goods Production Status and Year



Source: 2017 Longitudinal Business Database, Longitudinal Firm Trade Transactions Database, Economic Census, Special Inquiry data from the Census of Wholesale Trade.

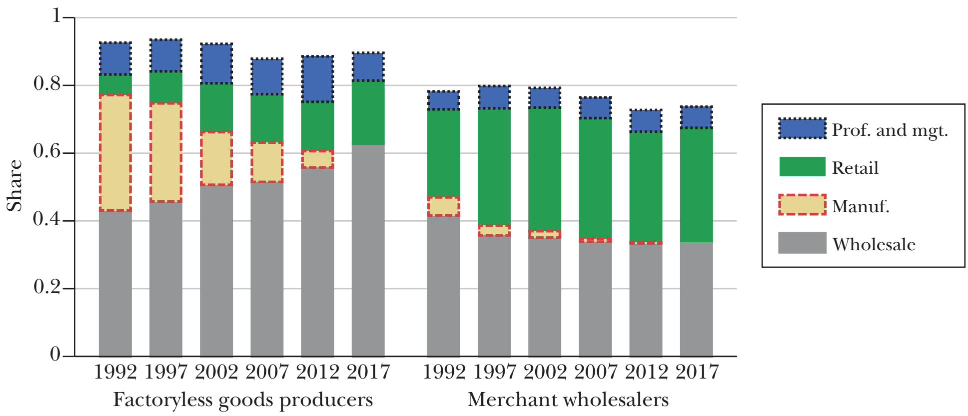
Note: Figure presents the ratio of imports to sales (left panel) and the share of firm imports from China over total imports (right panel). Factoryless goods producers are firms with one or more wholesale establishments that contract for manufacturing services in 2017. Merchant Wholesalers are firms with at least one establishment that reports not contracting for manufacturing and none that do. Sample is all firms with at least one wholesale establishment that responds to the 2017 questions on contract manufacturing services and with no manufacturing plants in 2017.

Perhaps the most striking message from Figure 3 is that the 2017 factoryless goods producers that were present in 1992 were considerably more involved in manufacturing. Indeed, these factoryless goods producers had over one-third of their workforce in manufacturing plants in 1992. The traditional wholesale firms in this sample (again, tracing them back from 2017 to 1992) have much lower manufacturing employment shares throughout and instead are more retail-intensive than factoryless goods producers. Their share of retail employment remains quite constant at about one-third over the last two decades. By contrast, the retail share of employment at factoryless goods producers doubles from 9.5 to 19 percent over that period.

Factoryless goods producers' share of employment in the knowledge-related categories of Professional and Management workers (NAICS 54–55) grows steadily from 9.3 to 13.4 percent in the period 1997–2012, though it then falls in 2017. It remains higher than the share of the comparison group, which hovers around 6 percent throughout. Factoryless goods producers share of such "knowledge workers" is thus not as high as the share of 0.19 at transnational manufacturers, but still substantially greater than the 0.03 share of purely domestic manufacturing firms.

These patterns suggest that the longer-lasting factoryless goods producers considered here were more directly involved in manufacturing in the past and have transitioned towards the pre- and postproduction stages as they increasingly import

Figure 3
Sector Employment Shares for 2017 Firms by Factoryless Goods Production Status
(*2017 Firms Present in 1992*)



Source: 2017 Longitudinal Business Database, Longitudinal Firm Trade Transactions Database, Economic Census, Special Inquiry data from the Census of Wholesale Trade.
Note: Figure plots firms' employment shares by sector. Factoryless goods producers are firms with one or more wholesale establishments that contract for manufacturing services in 2017. Merchant Wholesalers are firms with at least one establishment that reports not contracting for manufacturing and none that do. Sample is all firms with at least one wholesale establishment that responds to the 2017 questions on contract manufacturing services and with no manufacturing plants in 2017.

the goods they used to manufacture domestically. At least in these in these aggregate figures, however, the growth in imports trails the decline in manufacturing employment. While both firm types grow their total employment over the period, the factoryless goods producers grow from a weighted average of 40 workers per firm in 1992 to 62 workers by 2017 versus 44 to 99 for the comparison group (for details, see Appendix Table B.3).

The falling manufacturing employment shares at what become factoryless goods producers by 2017 suggest that these firms may have contributed to the decline in US manufacturing over the last several decades. Indeed, Fort et al. (2017) find that 75 percent of the decline in US manufacturing employment from 1997 to 2012 occurs in continuing firms, with two-thirds attributable to continuing firms' closure of manufacturing plants. Before concluding that this new organizational form has led to net declines in total employment, two caveats are in order. First, the set of continuing factoryless goods producers has grown its total employment over the period. Second, the information in Table 4 indicates that factoryless goods producers are relatively young, and examining those entrants' initial manufacturing employment suggests they may never have manufactured in-house.[8] Assessing the

[8] Appendix Figure B.1 shows that after 1997, the entering cohorts of eventual 2017 factoryless goods producers have lower manufacturing employment shares than the 1992 cohort (and than merchant wholesalers in the same cohort).

net employment effects of these former manufacturing firms and entering factoryless goods producers that never manufacture is an interesting question for future work, especially in light of their greater import intensity.

The higher employment shares of factoryless goods producers in knowledge-related activities resonates with the higher shares in these activities by multinational enterprises and with the prior findings that factoryless goods producers are more innovative. For example, Kamal (2023) finds that they have higher R&D expenditures, are more R&D-intensive, patent more, and own more trademarks than comparison service firms, using the 2011 Company Organization Survey and the 2012 Censuses of Wholesale and Services to identify factoryless goods producers. These patterns suggest that factoryless goods producers specialize in preproduction tasks, while outsourcing physical transformation tasks to other firms, often in other countries.

Sourcing Location Decisions by Factoryless Goods Producers

For additional evidence on the global orientation of factoryless goods producers, I calculate the extent to which they contract for manufacturing from domestic or foreign providers. An important caveat is that my data on firms' purchase locations are limited to firms that also responded to the question about whether they designed the products they outsourced. These firms cover 68 percent of the factoryless goods producers in my sample, and 75, 73, and 86 percent of their employment, sales, and imports, respectively. Table 5 presents the distribution of factoryless goods producers and their employment, sales, and imports based on whether their contracted manufacturing is performed in the United States, in foreign countries, or both. For each activity, the shares do not sum to one, due to the missing design and location information in the data.

The primary message from Table 5 is the dominance of contracting from foreign countries for factoryless goods producers: as the figure shows, at least 54 percent of factoryless goods producers and over 60 percent of their employment, sales, and imports are accounted for by firms that contract either exclusively or partly abroad. Indeed, a majority of all factoryless goods producers contract from foreign suppliers. For these firms, "outside the firm" also relates to "outside the country."

Reassuringly, Table 5 also shows that factoryless goods producers' imports are concentrated in firms that contract with suppliers in foreign countries. Firms that only contract with domestic suppliers account for just 2.9 percent of total imports by these firms. This high import share demonstrates that foreign purchases of contract manufacturing services and wholesale firms' merchandise imports are related.

The sourcing patterns of factoryless goods producers do not align well with standard trade models that rely on higher fixed costs of foreign sourcing to explain heterogeneous firms' selection into foreign markets. First, Table 5 indicates that firms with purely domestic or purely foreign sourcing are similarly sized (about 25 workers per firm) and pay comparable wages (about $70,000). Second, firms that rely exclusively on foreign suppliers are the most prevalent organizational form. Comparing these patterns to those for firms with manufacturing plants poses even bigger challenges for standard models: there seem to be far more domestic

Table 5
Factoryless Goods Producer Characteristics in 2017 by Supplier Location

	Share of Factoryless Good Producers				*Firm Characteristics*			
Supplier Locations	*Firms*	*Employment*	*Sales*	*Imports*	*Average Employment*	*Pay/ Employment*	*Sales/ Employment*	*Imports/ Sales*
Domestic Only	0.14	0.15	0.13	0.03	26	67	673	0.06
Foreign Only	0.42	0.40	0.41	0.63	25	70	794	0.38
Domestic & Foreign	0.12	0.21	0.19	0.21	45	99	728	0.27

Source: 2017 Longitudinal Business Database, Longitudinal Firm Trade Transactions Database, Economic Census, Special Inquiry data from the Census of Wholesale Trade.
Note: Table presents activity shares and weighted averages for factoryless goods producers and merchant wholesalers. Factoryless goods producers are firms that contract for manufacturing services from other firms. Sales in $1,000s. Sample in this table consists of the 25,200 factoryless goods producers that responded to the product design question.

manufacturing firms than factoryless goods producers, they are larger on average (about 46 workers per firm), their foreign sourcing (from in-house foreign plants or other firms) is rare, and the vast majority of those that offshore maintain significant domestic production.

Comparisons with Previous Findings on Factoryless Goods Producers

Past work on use of contract manufacturing services by manufacturing and wholesale firm must be compared to the present evidence with caution, given differences across data and samples, but several suggestive patterns emerge. First, factoryless goods producers seem much more outwardly oriented than manufacturing firms. Fort (2017) shows that approximately 30 percent of US manufacturing plants contract for manufacturing services from other firms in 2007, but among these firms, less than 7 percent do so from foreign suppliers. By contrast, one-quarter of all wholesale establishments that purchase contract manufacturing in 2007 also offshore (for details, see Appendix Figure C.3). This establishment comparison thus reinforces the conclusion that "out of the firm" and "out of the country" tend to go together.

Second, the prevalence of factoryless goods producers seems to have increased significantly from 2007 to 2017. While Bernard and Fort (2015) calculate that 12 percent of firms in their sample were factoryless goods producers in 2007, I use a reasonably similar calculation that implies approximately 27 percent of wholesale firms without manufacturing plants are factoryless goods producers by 2017.

Third, the foreign orientation of factoryless goods producers has also risen over this time period. In 2007, 3.7 percent of wholesale establishments contracted from foreign suppliers. Since 15 percent of wholesale establishments purchased any contract manufacturing services that year, about a quarter of the 2007 factoryless goods producers sourced from foreign suppliers. As shown in Table 5, this rate more than doubled by 2017, when at least 54 percent of factoryless goods producers source from foreign suppliers.

Finally, firms' use of contract manufacturing seems strongly related to their focus on innovation. In 2007, 45 percent of wholesale establishments that designed goods also purchased contract manufacturing services, compared to only 10 percent among nondesigners. Among the 2007 wholesale establishments that contracted for manufacturing, 29 percent that designed their own goods offshored, versus only 21 percent of establishments that did not design. Similarly, a majority of factoryless goods producers in 2017 design the goods they sell, and those that design are more likely to contract with foreign suppliers. These patterns are all consistent with the premise that factoryless goods producers tend to focus on preproduction manufacturing stages in the United States, while locating physical transformation tasks outside both the firm and the country.

The Interdependence between Integration and Location Decisions

US manufacturing firms have divided their manufacturing both across countries and across firms in ways that suggest interdependencies between these decisions. When firms perform physical transformation activities within the firm in foreign countries, the majority of their US sales and employment is also in manufacturing plants. However, when factoryless goods producers outsource physical transformation activities, they are increasingly likely to locate them in foreign countries. "In the firm" also coincides with a significant portion "in the country," while "outsourcing" maps to "offshoring."

The disproportionate focus on domestic innovation by both types of firm suggests that intellectual property is a key factor in their production processes. Some firms may specialize in design to increase their R&D efficiency, for example if there are increasing and convex costs to managerial scope, as in Lucas (1978). This type of specialization has been documented within manufacturing firms in response to increased foreign competition (Bernard, Redding, and Schott 2011; Mayer, Melitz, and Ottaviano 2014) and new offshoring opportunities (Bernard et al. 2023). From this view, factoryless goods producers just represent a more extreme form of specialization in pre- and postproduction tasks.

By contrast, other firms may improve research efficiency by using integrated manufacturing plants that are proximate to their headquarters and research centers. These plants may produce complex goods that are near the technology frontier or products for which manufacturing feeds back into research. While mature, stable products can be made far from the innovation hubs (Duranton and Puga 2001) and perhaps at arm's length, those at the technology frontier may benefit from face-to-face interactions with researchers. As an example, Boeing supports its "Design-to-Build" ethos by training the engineers at its South Carolina Propulsion plant as mechanics and tasking them with building parts to identify design opportunities to enhance production efficiency. Texas Instruments stresses synergies between their technology groups and manufacturing operations to ensure "manufacturability and cost efficiency." Although firms' US manufacturing plants are an average of several hundred miles away from their US R&D labs, firms tend to have at least one manufacturing plant co-located with R&D; moreover, those firms patent relatively more in those regions and time periods in which their manufacturing and knowledge establishments are co-located (Fort et al. 2020).

The importance of protecting intellectual property may also relate to firms' location and integration decisions. Firms for which theft is not a concern may be more likely both to outsource and offshore. This situation may arise either because their innovation is effectively excludable, as in the case of enforceable patents such as for pharmaceuticals, or because the product life cycle is sufficiently short, such as for fashion and phones. Indeed, US multinational enterprises disproportionately locate their in-house manufacturing affiliates in industries with long product lifecycles only in those countries with strong intellectual property protection (Bilir 2014).

Such industry differences, however, seem insufficient to explain the bifurcation in firms' integration and location decisions documented here. In 2007, Electrical Machinery and Equipment (HS 85) and Machine and Mechanical Appliances and Computers (HS 84) accounted for over 40 percent of imports of factoryless goods producers, compared to just 30 percent for the comparison group of firms (Bernard and Fort 2015). Multinational enterprises in these sectors comprise 17 percent of US employment by manufacturing multinationals in 2017, according to data from the Bureau of Economic Analysis. Thus, some US firms in computer and electronics and machinery maintain integrated manufacturing around the globe, while others outsource physical transformation tasks. For example, Apple and IBM both shed their personal computer manufacturing in 2004, but IBM continues to manufacture mainframes in the United States, while Apple ceased all in-house production tasks.

This bifurcation is evident even for a narrowly defined (and increasingly salient) product: the semiconductor chip. Texas Instruments and Qualcomm both sell chips, yet the former maintains integrated production, while the latter focuses solely on design and innovation. According to Kyle Flessner of Texas Instruments, "A core element of our strategy is to invest in increasing our internal manufacturing capacity—in wafer fabs and assembly-test sites we own—rather than relying only on external suppliers," whereas Qualcomm considers itself "a company of inventors with diverse skills and backgrounds."[9] These anecdotes suggest an important role for firm-level core competence and strategic focus in determining how firms organize their production across firms and countries.

Specialization in design may also arise when firms have differential access to capital and there are large fixed costs to manufacture—as for semiconductor manufacturing—such that one large contract manufacturer can potentially serve multiple designers more efficiently. Indeed, recent work finds that within-industry heterogeneity in firms' reliance on outsourcing is negatively correlated with their use of leverage, which is consistent with the presence of high fixed costs that finance (Moon and Phillips 2021). Because physical capital is often easier to sell and transfer than intangible capital, it is perhaps not surprising that lower-wage

[9] See https://news.ti.com/blog/2022/11/01/qa-how-ti-is-investing-in-manufacturing-capacity-to-support-growth-for-and https://www.qualcomm.com/company#about.

countries have specialized in the parts of the production that require that capital, while US firms increasingly specialize in intangibles.

Some existing models that study a firm's decision to outsource or offshore do feature firm heterogeneity, but only in one dimension. For example, in extensions of the Melitz (2003) model of international trade, a firm can open horizontal replicas of its domestic manufacturing plants abroad as an alternative to exports for serving foreign customers (Helpman, Melitz, and Yeaple 2004); or it can procure inputs in low- wage countries to lower production costs, in its own plants or from arm's-length suppliers (Antràs and Helpman 2004). These models capture salient features about US manufacturers—namely that exporters and importers are larger and more productive than nontraders, and that only the biggest and most productive own foreign plants (Bernard et al. 2007; 2018). However, there is no interaction between firms' location and integration decisions, and the fixed-cost ordering that can explain multinationals' behavior does not rationalize the patterns of factoryless good producers documented here.

The divergence in firm types documented here thus calls for a new framework to analyze both of the firm's boundary decisions jointly. If integration and location decisions are interdependent, then changes in trade policy will not only affect the location of production, but also influence the scope of firms' integrated activities. Similarly, changes in monitoring or other technologies that reduce contracting frictions (for example, Baker and Hubbard 2004) may also change production location decisions. Such interactions create new challenges and opportunities for assessing the effects of changing trade costs. At a broader level, they relate to insights from Holmstrom and Milgrom (1991), who emphasize the role of complementarities across tasks in optimal job design and firm structure.

The interdependence in firms' outsourcing and offshoring decisions also has important implications for empirical work. A common approach to analyze the effects of trade is to regress industry-level changes in outcomes (such as employment) on instrumented trade flows. When reallocation occurs across firms and industries, however, this method may paint an incomplete picture. For example, this method would capture Apple's exit from US manufacturing, but miss its related growth in innovation and retail sectors. Recent work documents decreased US patenting by public manufacturing firms in response to increased Chinese imports (Autor et al. 2020). Yet results from a new dataset of US patents from 1977 to 2016 by all firms in the United States indicate that the share of patents granted to manufacturing firms has fallen dramatically, from 91 to 54 percent between 1977 to 2016, while patents by former manufacturing firms have grown steadily, especially for firms that grew their Chinese imports after 2007 (Fort et al. 2020).

Apple's shift from manufacturing to design also highlights the importance of distinguishing global value chain trade from import competition. It is now well established that trade flows from fragmented production have potentially different effects from the more standard "wine-for-cloth" exchange of final goods. For example, Feenstra and Hanson (1999) show that US input trade with lower skill countries can increase the demand for skilled workers within an industry as domestic producers focus on a subset of higher skill tasks and sourced lower skilled tasks from abroad.

Ding et al. (2022) document precisely this reallocation in response to cheaper inputs from China. They show that US firms that rely on inputs for which China gained market share in Europe increase both the shares and levels of their nonmanufacturing employment. However, input trade misses final-good trade by both transnational manufacturers and factoryless goods producers. Yesterday's efforts to measure global value chains and fragmentation using trade in intermediate inputs simply do not capture today's reality in which US firms sell final goods manufactured abroad but designed, distributed, and marketed using domestic labor and ideas.

Conclusion

US manufacturers are connected to global value chains in multiple ways. Some firms have opened in-house manufacturing plants in foreign countries, yet maintain domestic manufacturing as a primary domestic activity. Other firms both outsource and offshore the traditional manufacturing stages, yet remain involved in the broader production process by designing and marketing their goods and coordinating across their arm's-length suppliers. Despite their contrasting organizational forms, both transnational in-house manufacturers and factoryless goods producers hire disproportionately more knowledge workers in the United States. They also spend more on R&D and receive more patent grants. These patterns highlight the need for new trade models in which low-wage manufacturing locations enable the entry of more ideas by firms that specialize in domestic innovation.

Understanding how US firms leverage their domestic knowledge creation across countries is also necessary for producing reliable estimates of GDP, value-added, and productivity. When US firms sell their products directly to foreign customers from their foreign suppliers or plants, those goods never cross into US commercial space. The ensuing profits are counted in US GNP, but the value added by US designers and software engineers may be excluded from GDP. Guvenen et al. (2022) estimate that US multinationals shift between \$150 to \$200 billion per year in profits using their foreign affiliates, with most of the shifting in R&D-intensive industries and firms. This issue may be most severe for factoryless goods producers, because they are fully specialized in the pre- and postproduction stages that add considerable value to the final product, yet are not readily observable in aggregate statistics. As an example, Bayard, Byrne, and Smith (2015) use Apple's global revenue from iPad sales reported in its 2011 annual report to calculate that (under plausible assumptions about the gross margins on sales of different products) value added in the US Computer industry would have been approximately \$6 billion higher if Apple's value-added from its iPads were included, roughly offsetting the decline in domestic computer manufacturing that year.

The results in this paper thus also point to the need for statistical agencies to improve the available data for studying the fragmentation of knowledge production and manufacturing. First, statistical agencies could collect sales, inputs, imports, and exports using the same classification systems, which would allow for more accurate assessments of what firms buy, sell, import, export, and produce. Second, collection of these elements could be expanded across sectors, perhaps using techniques that allow for automatic recording and transmission of transactions, rather than the traditional survey approach. Third, a flag could be added to the US Customs import form asking whether the goods were produced by contract manufacturers for the importer.[10] Finally, data on firms' technology use would facilitate studies about the ways in which cross-county teams form and operate.

Such expansions of data collection are crucial for a complete picture of global production sharing and accurate assessments of US supply-chain risk. Past work cleverly leverages input-output tables to document production sharing across countries (Hummels, Ishii, and Yi 2001; Johnson and Noguera 2012), but those metrics will remain incomplete until the underlying data sources link the value added by firms in one country to the output of different firms in other countries and across sectors. These links are essential to analyze the costs and benefits of potential changes in trade costs, such as the recent proposal by the US National Security Advisor, Jake Sullivan (2023), to protect US technology "with a small yard and high fence." Such proposals may upset the current balance between domestic innovation and foreign physical transformation. Factoryless goods producers may be particularly susceptible, as they cannot relocate their suppliers' unilaterally.

Perhaps the most exciting direction for future work is to study how foreign outsourcing of physical transformation tasks affects the creation and diffusion of knowledge. Research on foreign direct investment studies these transfers explicitly (Javorcik 2004; Ramondo and Rodríguez-Clare 2013; Arkolakis et al. 2018). Because excluding knowledge from rivals is one motive for integration (Ding et al. 2022), the largest flows may occur when arm's-length relationships form. Factoryless goods production thus represents a new form of global manufacturing with the potential to spread ideas around the world.

[10] The administrative value-added tax data collected in a number of countries might also be used to distinguish factoryless goods producers from traditional service firms. For instance, if countries with those data could collect information on sourcing for the firm's own goods by means of contract manufacturers, we could assess whether factoryless goods producers tend to have more long-lasting and concentrated relationships with their suppliers. Improvements in text-based algorithms that allow for detailed concordances across classification systems may be a short-term solution to the concordance challenges from using US data.

■ *Any views expressed are those of the author and not those of the US Census Bureau, the Bureau of Economic Analysis, the National Bureau of Economic Research, or the Centre for Economic Policy Research. The Census Bureau and the Bureau of Economic Analysis have reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 1975 (CBDRB-FY23-P1975-R10185), CES Project 1530 (release on 7/15/2019), and CES Project 6907751 (BEA-FY23-P6907751-R2 and CBDRB-FY23-CED006-0010). Jack Liang provided exceptional research assistance. I thank Pol Antràs, Jim Fetzer, Fariha Kamal, Peter Schott, and Chad Syverson for very helpful comments. I also thank the editors, Erik Hurst, Nina Pavcnik, Heidi Williams, and especially Timothy Taylor for invaluable insights, comments, and suggestions.*

References

Antràs, Pol, and Elhanan Helpman. 2004. "Global Sourcing." *Journal of Political Economy* 112 (3): 552–80.

Antràs, Pol, Evgenii Fadeev, Teresa C. Fort, and Felix Tintelnot. 2023. "Exporting, Global Sourcing, and Multinational Activity: Theory and Evidence from the United States." NBER Working Paper 30450.

Arkolakis, Costas, Natalia Ramondo, Andrés Rodríguez-Clare, and Stephen Yeaple. 2018. "Innovation and Production in the Global Economy." *American Economic Review* 108 (8): 2128–73.

Autor, David, David Dorn, Gordon H. Hanson, Gary Pisano, and Pian Shu. 2020. "Foreign Competition and Domestic Innovation: Evidence from US Patents." *American Economic Review: Insights* 2 (3): 357–74.

Baker, George P., and Thomas N. Hubbard. 2004. "Contractibility and Asset Ownership: On-Board Computers and Governance in U.S. Trucking." *Quarterly Journal of Economics* 119 (4): 1443–79.

Bayard, Kimberly, David Byrne, and Dominic Smith. 2015. "The Scope of U.S. 'Factoryless Manufacturing.'" In *Measuring Globalization: Better Trade Statistics for Better Trade Policy*, Vol. 2, edited by Susan N. Houseman and Michael Mandel, 81-120. Kalamazoo, MI: W.E. Upjohn Institute.

Berman, Eli, John Bound, and Zvi Griliches. 1994. "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufactures." *Quarterly Journal of Economics* 109 (2): 367–97.

Bernard, Andrew B., and Teresa C. Fort. 2015. "Factoryless Goods Producing Firms." *American Economic Review: Papers & Proceedings* 105 (5): 518–23.

Bernard, Andrew B., and Teresa C. Fort. 2017. "Factoryless Goods Producers in the USA." In "The Factory-Free Economy," edited by Lionel Fontagne and Ann Harrison, 113–44. Oxford, UK: Oxford University Press.

Bernard, Andrew B., J. Bradford Jensen, and Peter K. Schott. 2009. "Importers, Exporters, and Multinationals: A Portrait of Firms in the U.S. That Trade Goods." In *Producer Dynamics: New Evidence from Micro Data*, edited by Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, 513–52. Chicago: University of Chicago Press.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2011. "Multiproduct Firms and Trade Liberalization." *Quarterly Journal of Economics* 126 (3): 1271–1318.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2007. "Firms in International Trade." *Journal of Economic Perspectives* 21 (3): 105–30.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2018. "Global Firms." *Journal of Economic Literature* 56 (2): 565–619.

Bernard, Andrew B., Teresa C. Fort, Valerie Smeets, and Frederic Warzynski. 2023. "Heterogenous Globalization: Offshoring and Reorganization." NBER Working Paper 26854.

Bernard, Andrew B., Emmanuel Dhyne, Glenn Magerman, Kalina Manova, and Andreas Moxnes. 2022. "The Origins of Firm Heterogeneity: A Production Network Approach." *Journal of Political Economy* 130 (7): 1765–1804.

Bilir, L. Kamran. 2014. "Patent Laws, Product Life-Cycle Lengths, and Multinational Activity." *American Economic Review* 104 (7): 1979–2013.

Boehm, Christoph E., Aaron Flaaen, and Nitya Pandalai-Nayar. 2020. "Multinationals, Offshoring, and the Decline of U.S. Manufacturing." *Journal of International Economics* 127: 103391.

Braguinsky, Serguey, Atsushi Ohyama, Tetsuji Okazaki, and Chad Syverson. 2021. "Production Innovation, Product Diversification, and Firm Growth: Evidence from Japan's Early Industrialization." *American Economic Review* 111 (12): 3795–3826.

Chow, Melissa C., Teresa C. Fort, Christopher Goetz, Nathan Goldschlag, James Lawrence, Elisabeth Ruth Perlman, Martha Stinson, and T. Kirk White. 2021. "Redesigning the Longitudinal Business Database." NBER Working Paper 28839.

Ding, Xiang, Teresa C. Fort, Stephen J. Redding, and Peter K. Schott. 2022. "Structural Change within versus across Firms: Evidence from the United States." NBER Working Paper 30127.

Doherty, Maureen. 2015. "Reflecting Factoryless Goods Production in the U.S. Statistical System." In *Measuring Globalization: Better Trade Statistics for Better Trade Policy*, Vol. 2, edited by Susan N. Houseman and Michael Mandel, 13–43. Kalamazoo, MI: W.E. Upjohn Institute.

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

Feenstra, Robert C. 2010. *Offshoring in the Global Economy: Microeconomic Structure and Macroeconomic Implications*. Cambridge, MA: MIT Press.

Feenstra, Robert C., and Gordon H. Hanson. 1999. "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the United States, 1972-1990." *Quarterly Journal of Economics* 114 (3): 907–40.

Fort, Teresa C. 2017. "Technology and Production Fragmentation: Domestic versus Foreign Sourcing." *Review of Economic Studies* 84 (2): 650–87.

Fort, Teresa C. 2023. "Replication data for: The Changing Firm and Country Boundaries of US Manufacturers in Global Value Chains." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E192222V1.

Fort, Teresa C., Wolfgang Keller, Peter K. Schott, and Stephen Yeaple. 2020. "Colocation of Production and Innovation: Evidence from the United States." Unpublished.

Fort, Teresa C., and Shawn D. Klimek. 2018. "The Effects of Industry Classification Changes on US Employment Composition." Center for Economic Studies Working Paper 18-28.

Fort, Teresa C., Justin R. Pierce, and Peter K. Schott. 2017. "New Perspectives on the Decline of US Manufacturing Employment." 32 (2): 47–72.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection of Productivity or Profitability?" *American Economic Review* 98 (1): 394–425.

Garetto, Stefania, Lindsay Oldenski, and Natalia Ramondo. 2019. "Multinational Expansion in Time and Space," NBER Working Paper 25804.

Grossman, Gene M., and Elhanan Helpman. 1991. "Quality Ladders in the Theory of Growth." *Review of Economic Studies* 58 (1): 43–61.

Guvenen, Fatih, Raymond J. Mataloni Jr., Dylan G. Rassier, and Kim J. Ruhl. 2022. "Offshore Profit Shifting and Aggregate Measurement: Balance of Payments, Foreign Investment, Productivity, and the Labor Share." *American Economic Review* 112 (6): 1848–84.

Helpman, Elhanan, Marc J. Melitz, and Stephen R. Yeaple. 2004. "Export versus FDI with Heterogenous Firms." *American Economic Review* 94 (1): 300–316.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7 (S): 24–52.

Hummels, David, Jun Ishii, and Kei-Mu Yi. 2001. "The Nature and Growth of Vertical Specialization in World Trade." *Journal of International Economics* 54 (1): 75–96.

Jarmin, Ron S., and Javier Miranda. 2002. "The Longitudinal Business Database," Center for Economic Studies Working Paper 02-17.

Javorcik, Beata Smarzynska. 2004. "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages." *American Economic Review* 94 (3): 605–27.

Johnson, Robert C., and Guillermo Noguera. 2012. "Accounting for Intermediates: Production Sharing and Trade in Value Added." *Journal of International Economics* 86 (2): 224–36.

Jones, Ronald W., and Henryk Kierzkowski. 2001. "Globalization and the Consequences of International Fragmentation." In *Money, Capital Mobility, and Trade: Essays in Honor of Robert A. Mundell*, edited by Guillermo A. Calvo, Rudi Dornbusch, and Maurice Obstfeld, 365–83. Cambridge, MA: MIT Press.

Kamal, Fariha. 2023. "A Portrait of US Factoryless Goods Producers." In *Challenges of Globalization in the Measurement of National Accounts*, edited by Nadim Ahmad, Brent Moulton, J. David Richardson, and Peter Van De Ven, 425–50. Chicago: University of Chicago Press.

Kamal, Fariha, and Wei Ouyang. 2020. "Identifying US Merchandise Traders: Integrating Customs Transactions with Business Administrative Data." Center for Economic Studies Working Paper 20-28.

Kamal, Fariha, Jessica McCloskey, and Wei Ouyang. 2022. "Multinational Firms in the U.S. Economy: Insights from Newly Integrated Microdata." Center for Economic Studies Working Paper 22-39.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy. 2021. "Measuring Technological Innovation over the Long Run." *American Economic Review: Insights* 3 (3): 303–20.

Lucas, Robert E., Jr. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9 (2): 508–23.

Mayer, Thierry, Marc J. Melitz, and Gianmarco I. P. Ottaviano. 2014. "Market Size, Competition, and the Product Mix of Exporters." *American Economic Review* 104 (2): 495–536.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.

Moon, S. Katie, and Gordon Phillips. 2021. "Outsourcing through Purchase Contracts and Firm Capital Structure." *Management Science* 67 (1): 363–87.

Office of Management and Budget. 2011. "North American Industry Classification System; Revision for 2012." *Federal Register* 76 (159): 51240–43.

Ramondo, Natalia, Veronica Rappoport, and Kim J. Ruhl. 2016. "Intrafirm Trade and Vertical Fragmentation in U.S. Multinational Corporations." *Journal of International Economics* 98: 51–59.

Ramondo, Natalia, and Andrés Rodríguez-Clare. 2013. "Trade, Multinational Production, and the Gains from Openness." *Journal of Political Economy* 121 (2): 273–332.

Rodríguez-Clare, Andrés. 2010. "Offshoring in a Ricardian World." *American Economic Journal: Macroeconomics* 2 (2): 227–58.

Sullivan, Jake. 2023. "Remarks by National Security Advisor Jake Sullivan on Renewing American Economic Leadership at the Brookings Institution." Speech, Brookings Institution, Washington, DC, April 27, 2023. https://www.whitehouse.gov/briefing-room/speeches-remarks/2023/04/27/remarks-by-national-security-advisor-jake-sullivan-on-renewing-american-economic-leadership-at-the-brookings-institution/.

Journal of Economic Perspectives—Volume 37, Number 3—Summer 2023—Pages 59–86

Global Value Chains in Developing Countries: A Relational Perspective from Coffee and Garments

Laura Boudreau, Julia Cajal-Grossi, and Rocco Macchiavello

On the morning of April 24, 2013, the garment factories of the Rana Plaza building in the Savar industrial area in Dhaka, Bangladesh, started their generators to make up for the recent power outages. Then the building collapsed, causing the death of 1,134 people and injuring 2,500 more. The building had compromised structural integrity: several floors had been added without a building permit. The Rana Plaza disaster—one of the deadliest industrial disasters in history—brought the working conditions in the garment sector in Bangladesh into the global spotlight. This and several other industrial disasters contrasted with phenomenal success along other measures: the Bangladesh garment sector, virtually nonexistent in the early 1980s, had averaged an annual growth rate in recent decades above 10 percent, accounted for 70–80 percent of Bangladesh's exports, and employed nearly 4 million workers, mostly women, in a country in which women had traditionally not worked outside the home.

At about the same time, Nespresso was attempting to revitalize South Sudan's export-oriented coffee industry in partnership with a nongovernmental organization called TechnoServe. The Republic of South Sudan emerged from decades of civil war as the world's youngest country on July 9, 2011. The civil war had left a legacy of abysmal maternal mortality rates and illiteracy, almost no functioning

■ *Laura Boudreau is Assistant Professor of Economics, Graduate School of Business, Columbia University, New York City, New York. Julia Cajal-Grossi is Assistant Professor of Economics, Geneva Graduate Institute (IHEID), Geneva, Switzerland. Rocco Macchiavello is Professor of Management, London School of Economics, London, United Kingdom. Their email addresses are l.boudreau@columbia.edu, julia.cajal@graduateinstitute.ch, and r.macchiavello@lse.ac.uk, respectively.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.3.59.

infrastructure, and a ravaged economy. The country's natural resources—mostly oil—attracted attention and gave hope that these resources could be mobilized to revive the economy. However, South Sudan had a long coffee tradition before the civil war destroyed the country's production capacity. Nespresso and TechnoServe trained thousands of farmers and established wet mills to meet export-grade requirements. Nespresso, which purchased all of the country's exports, finally launched the Limited Edition Grand Cru Suluja ti South Sudan in the US and European markets in 2016.

These examples highlight the complex governance issues, as well as the potential both for benefits and for costs, that arise in the global value chains that now account for almost half of global trade (World Bank 2020). There is somewhat widespread consensus, among policymakers and academics alike, that global value chains taken as a whole have helped developing countries grow and lifted many out of poverty. But as these examples suggest, developing countries differ from more advanced economies in important ways—such as an often weaker institutional environment, poorer state capacity for enforcing regulations, and persistent political instability, among others. Their participation in global supply chains raises contentious issues. For example, many observers believe that the market power that large international buyers wield in many supply chains results in unfairly low prices paid to workers and producers and in undesirably poor working conditions and quality standards (Gresser and Tickell 2002; Locke 2013).

In this essay, we will focus on the coffee and garment supply chains, which are classic examples of buyer-driven (Gereffi et al. 2001) global value chains. In these chains, production takes place in developing countries, and buyers from higher-income countries influence standards and terms of trade, making the contentious issues mentioned just above particularly salient. We begin with an overview of these supply chains. Alongside buyers' market power, we then emphasize that these supply chains operate in contexts where complete contracts are not possible due to a range of issues from measuring quality to unexpected shocks. These contracting problems are often compounded by the distinctive institutional features of developing countries.

To address these contracting problems and improve market outcomes, a common approach among participants in the coffee and garment industry is to rely on long-term relationships between buyers and sellers. Thus, our discussion emphasizes a relational view of trade, as described by Antrás (2020) in his review of the conceptual aspects in the study of global value chain. At the export gate, we emphasize the importance of long-term supply relationships between exporters and foreign buyers related to issues like quality, financing terms, and reliability. Beyond the export gate, the importance of relationships manifests itself in the interlinked transactions between smallholder farmers, first-stage processors, exporters in the case of coffee (and other agricultural chains), and in the quality of industrial relations between exporters and workers in the apparel sector. Finally, we discuss how long-term supply relationships at the export gate can be leveraged to improve relationships in the domestic part of the chain and address sustainability challenges, including environmental ones. For producers in developing countries, participation in long-term supply relationships can promote upgrading in product quality and management practices (for a review, see Verhoogen forthcoming) and—increasingly—in social and environmental standards.

This perspective on global value chains requires going beyond standard datasets about quantities and prices as recorded at national borders, which lack the detail necessary to understand how long-term supply relationships function. Instead, a relational perspective requires contextual and detailed knowledge. In this spirit, we focus on lessons we have learned over more than a decade working in partnership with a variety of stakeholders and several coauthors in the coffee and the garment chains in developing countries. When does monopsony power—like that held by Nespresso in South Sudan—depress prices paid to farmers and efficiency? When does it enable investments in otherwise prohibitively risky contexts? When does monopsony power of large garment buyers cause garment producers to take short-cuts that compromise on workers' safety and well-being? When does it promote fairer working conditions? Addressing these kinds of questions requires looking beyond measures of border trade, and instead understanding how long-term relationships will sometimes be able to address contractual frictions.

We argue that understanding market power and relationships—and how they relate to each other and at the different stages of the chain—is necessary to foster equitable and sustainable participation of developing countries in global value chains. Market power typically generates distortions relative to a first-best benchmark and inequitable distributional outcomes. The contracting problems highlighted above, however, suggest that first-best is not the relevant benchmark in most practical settings. In both garments and coffee, we show that proxies for market power and for relationships are positively correlated with each other both at the export gate and in the domestic portion of the chain. This suggests that the welfare consequences of market power cannot be assessed exclusively in terms of prices as, due to contracting problems, many other aspects of the transaction are important. Furthermore, a certain degree of market power might be needed to sustain beneficial long-term relationships. We need to know more about how to build and maintain well-functioning relationships that enable a more equitable participation in global value chains.

We make no pretense that the coffee and garment supply chains are representative of all supply chains. Indeed, as we shall explain, even these two supply chains are organized quite differently across developing countries. However, we do believe that the themes we explore using these two chains as laboratories—contracting problems, market structure, sustainability, and the importance of long-term supply relationships—are relevant in many other international trade contexts.

A Bird's-Eye View of Two Value Chains: Coffee and Garments

We begin with a succinct description of the coffee and garments global value chains, focusing on the export-oriented links of the chain in developing countries.

As these two sectors illustrate, the supply chain can look quite different across contexts.

Coffee

Coffee is produced in about 50 countries in the "coffee belt"—between 25° latitude North and South of the equator—and is the main source of livelihood for an estimated 25 million smallholders. Linking these producers to global value chains can potentially increase their incomes and alleviate poverty. Besides its intrinsic interest, the coffee chain is characterized by buyers' market power over producers (Watkins and Fowler 2002), but also, as we shall see in a moment, by several contracting challenges.

The top panel of Figure 1 illustrates the coffee supply chain in developing countries. The coffee cherry is the fruit of the coffee tree. After harvest, the bean inside the cherries is separated from the pulp, dried (and called "parchment coffee" at this stage), and then hulled and sorted to obtain green coffee. Coffee-producing countries, and even regions, differ in the extent to which farmers are involved in pulping, washing, and drying. In some countries (like Colombia), these activities are mostly undertaken by farmers that sell "parchment coffee" to intermediaries who then take it to exporters. In other countries (like Costa Rica), farmers sell coffee cherries to mills (washing stations) and deliver it to exporters. Most green coffee from producing countries is exported to traders or directly to roasters, before reaching retailers.

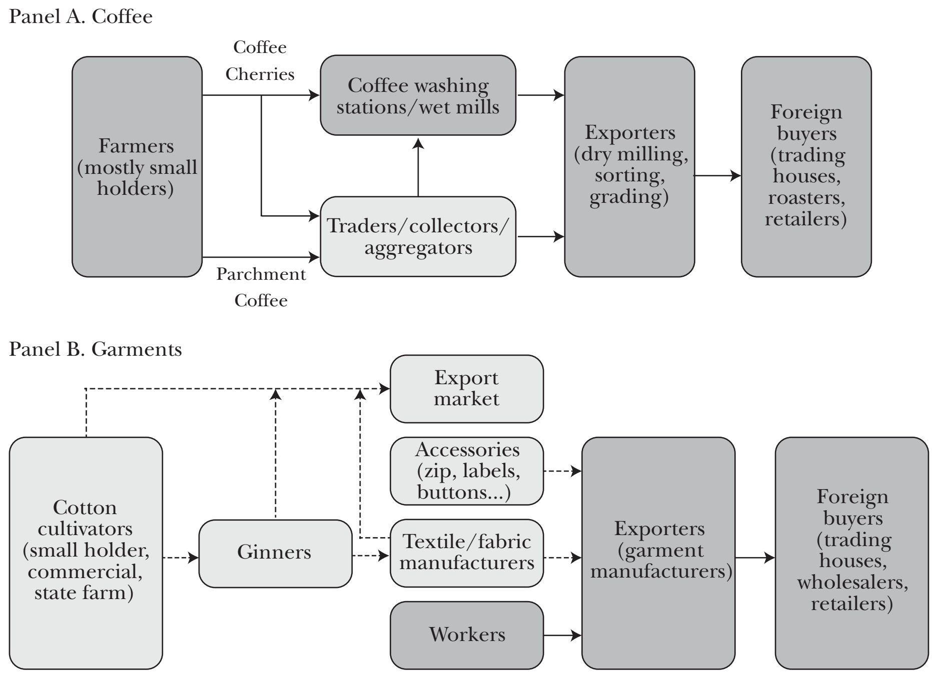
The cost of roasted coffee accounts for only 5–10 percent of the final price for a cup of coffee paid by consumers in high-income countries. Rent for the premises, labor costs, and taxes account for the vast majority of the final retail price. In turn, only a small fraction (10–15 percent) of the price paid by the retailer to the roasters reaches the farmer, with the rest being absorbed by roasters', traders' and processors' costs and margins along the chain (ICO 2020).

Garments

The garments sector makes intensive use of unskilled labor, in part because it has proven difficult to automate large parts of the production process. Thus, developing countries with abundant labor and relatively low wages have a comparative advantage, and the garment industry has in the past played a critical role in the early phases of export-oriented industrialization (Akamatsu 1962; Baldwin and Martin 1999), and most recently in East Asia (for example, Gereffi 1999). Rapidly falling trade barriers, like the phasing out of the international import quotas for apparel under the Multi-fiber Agreement in the 1990s and early 2000s, have induced a rapid expansion of garment production in developing countries. As wages in China's coastal areas rise, a large share of the world's garment production seems likely to relocate to poorer regions in South Asia and possibly Africa.

From a development perspective, two aspects of garment production are noteworthy. First, exported garments are typically produced in large manufacturing plants organized along production lines, which means that garments provide an

Figure 1
Organization of the Supply Chains of Coffee and Garments



Source: Authors' stylized representation of the supply chains of coffee (panel A) and garments (panel B).
Notes: The boxes depict different actors by their role in the supply chain. Boxes with gray shading are the focus of this article. Arrows reflect the flow of goods between actors, and those with solid lines are discussed in this article.

important training ground for modern management systems in developing countries (Bloom and Van Reenen 2010). Furthermore, it has been argued that women have a comparative advantage in sewing and stitching, the most labor-intensive steps in the production of garments. Garments have thus been an engine for women's emancipation.

The bottom panel of Figure 1 illustrates the export-oriented garment supply chain in developing countries. The production of garments is the last step in a process that starts in cotton fields, passes through textile companies that process yarn into cloth, and then brings together accessories and other inputs before workers cut, stitch, and package garments for exports. On the other side of the export gate, foreign buyers are typically brands, wholesalers, and retailers in high-income countries. Fabric is the main material input in the production of garments, accounting for 70–80 percent of the cost of a standard piece of garment as it leaves the factory gate. Labor, mostly employed in the sewing and stitching of garments, accounts for approximately 20 percent of the costs. Among the largest developing countries exporting garments, China, India, and Pakistan are also large producers of cotton and fabric, while Vietnam and Bangladesh mostly rely on imports.

At the Export Gate

In this section, we discuss the nature of the exchange between exporters and buyers in the two chains. The quantity of the product is observable and reported in statistics. Customs data is commonly reported according to the Harmonized Commodity Description and Coding System (see the website of the International Trade Administration at https://www.trade.gov/harmonized-system-hs-codes) and typically referred to as HS. However, the quality of products is harder to observe at the border. Conversations with stakeholders in coffee and garments furthermore reveal the importance of the two-way provision of services, qualities, and "promises" about how parties expect to trade in the future. These aspects of the exchange are almost never recorded in administrative datasets.

What Quality Is Being Traded?

Customs data provides only so much information about the product that is traded. Coffee is covered in the HS heading 0901: coffee, whether or not roasted or decaffeinated; coffee husks and skins; coffee substitutes containing coffee in any proportion. Within the heading, there are only six HS codes at six digits of disaggregation (HS6). Among these, over 90 percent of exports from producing countries is in 090111 (coffee, not roasted, not decaffeinated). Garments include products made from knitted or crocheted fabrics (HS chapter 61) and clothing made from woven fabrics (HS chapter 62). These two chapters span approximately 300 six-digit product codes. For example, 610510 is men's or boys' shirts, knitted or crocheted, of cotton.

However, buyers and sellers typically reach a common understanding of detailed quality specifications that goes beyond the product codes. For example, green coffee after milling is graded and classified for export with the aim of producing lots that meet certain quality criteria. There is no universally accepted quality grading system—each producing country has developed its own classification—but classifications that have been widely accepted across the industry support price negotiations between parties. For example, contracts are typically explicit about the coffee grade (the size of the coffee beans), the maximum rate of defects, and certifications. In garments, delineating quality parameters often entails the exchange of samples and post-shipment checks. Parties agree on an allowance for a certain percentage of defective garments.

Our conversations with stakeholders in both industries suggest that quality is in general observable, in the sense that buyers can observe it after seeing it, but not contractible, in the sense that it would be difficult for a court to adjudicate a contractual dispute over quality in a cost-effective manner. Asymmetric information over quality plays a bigger role for complex products, or those that must satisfy sanitary and phytosanitary (being free of crop disease) requirements. The performance of inspection authorities can then influence a country's collective reputation. For example, Bai, Gazze, and Wang (2021) provide a fascinating account of how a contamination scandal at certain producers affected all of Chinese dairy exports. An increasingly important dimension of quality relates to sustainability: consumers care not only about the final output, but also about how the product is manufactured or sourced. These dimensions of quality are harder to observe and pose more severe information and contracting problems. We return to these topics later in the paper.

How Is Trade Financed?

Financing terms are another key aspect rarely observable in standard datasets. On one side, an exporter can require the buyer to pay for goods before they are shipped. Alternatively, an exporter can extend trade credit to the importer, accepting payment after the goods have arrived at the destination. In the former case, the buyer incurs a risk of default if the exporter does not deliver; in the latter, the exporter bears the risk of nonpayment. Financial markets in developing countries are generally less developed, and firms are more likely to be credit-constrained (Banerjee and Munshi 2004; Banerjee and Duflo 2014). Moreover, international transactions involve longer delivery times (thus increasing working capital requirements) and parties located in different jurisdictions (increasing the costs of debt recovery). Antràs and Foley (2015) provides a comprehensive analysis of how exporters and importers navigate this tradeoff.

Financing the working capital required to produce for export is a first-order concern in both coffee and garments. The cost of coffee cherries sourced during harvest amounts to 70–80 percent of an exporter's seasonal revenues. Due to volatile weather and prices, lending to coffee exporters is risky, and banks tend to steer away from the sector, despite widespread state-sponsored support schemes available in many countries. Exporters thus commonly receive prefinancing from buyers. The buyer may advance funds necessary to finance 40–60 percent of the cost of cherries needed to deliver the agreed volume of coffee. In some cases, lenders accept the contract with the buyer as a form of collateral. In either case, the relationship with the foreign buyer is a source of collateral for the exporter. Blouin and Macchiavello (2019) analyze detailed data from one such scheme. They find that, even with such contractual arrangements in place, many coffee exporters are credit-constrained and process too little coffee, possibly depressing prices paid to farmers.

Many countries export garments through the cut-make-trim (or cut-make-package) system, in which the foreign buyer provides all the material inputs to the exporter, who finances the labor. Given that fabric and materials jointly account for more than 70 percent of the variable costs of production, this system drastically reduces working capital requirements. However, the system also limits the potential for the exporter of capturing a higher share of the value added by entering additional steps of production (like sourcing, logistics, and so on). Financial frictions thus impact the organization of production and the potential for upgrading (Manova and Yu 2016). Some observers have credited the success of the Bangladesh garment industry to the system of "back-to-back" letters of credit enabled by the central bank. Under this credit facility, exporters import material from abroad using a letter of credit from the buyer as a guarantee. This has allowed Bangladeshi exporters to control more functions and capture a higher share of value addition.

Other Dimensions: Reliability, Flexible Supply, and Demand Assurance

Reliability of supply (the supplier's ability to deliver orders with no delay and according to agreed-upon specifications) is the most recurrent aspect mentioned by buyers in conversations about suppliers' performance. However, the reliability of a given supplier can be difficult to assess—which makes a supplier's reputation for reliability a valuable asset. Standard datasets record the timing of the trade that took place, not its discrepancy from what parties had agreed upon. Reliability, let alone a reputation for it, is thus hard to observe in data normally available to researchers.

Macchiavello and Morjaria (2015) provides a vivid illustration of the importance of maintaining a reputation for reliability using the Kenya flower sector as a case study. Ethnic-based violence erupted in several parts of the country in early 2008 following the heavily contested presidential election. Due to workers' shortages, many exporters could not harvest flowers in their greenhouses (Ksoll, Macchiavello, and Morjaria 2022). Although exporters exerted costly efforts in order to continue to reliably supply their long-term buyers, many were not able to honor agreements with all their customers and needed to choose which ones to prioritize. Because the behavior of an exporter potentially signals future reliability to customers, exporters tend to prioritize their most established customers. Up to a certain point, that is: the exporter has nothing "left to prove" to buyers where the relationship is already strong. In other words, reputation implies an inverted-U shape relationship between reliability during the shock and the exporter's previous experience with the buyer—a prediction well-supported by the data.

Flexibility refers to the supplier's ability to accelerate production, allocate additional capacity, or accommodate changes in design, all at short notice. Flexibility is especially important when demand is hard to predict. Buyers partially address their own need for flexibility and supply assurance by maintaining some production "closer" to where goods are sold, even if at a higher cost, or by using more expensive suppliers less regularly. For example, Gap maintains a relatively small number of suppliers in Mexico and Central America, and Inditex (the owner of Zara) does so in Spain, Portugal, and Morocco, despite higher labor costs compared to Asia.

Conversely, stable and predictable demand helps suppliers optimize capacity utilization. In coffee, much trade happens through forward contracts in which exporters commit to deliver, and buyers to accept, coffee at a later date. These contracts are often agreed upon before the beginning of the harvest season and provide stability to both parties. Macchiavello and Miquel-Florensa (2017) document that sales agreed very early or very late in the season fetch up to 5–10 percent lower prices due to these demand assurance and inventory risk concerns. If a supplier decided not to deliver the promised coffee, perhaps to take advantage of more profitable market conditions at delivery, there would be little that a buyer could do to be compensated. In garments, buyers often book production capacity not just during seasonal peaks, but also during the less busy periods, enabling exporters to utilize capacity more efficiently. Again, if a buyer was to renege on that promise, perhaps because a cheaper supplier has been found, there would be little that an exporter could do to claim compensation.

For all of these reasons, the information recorded in standard datasets provides a limited characterization of what is traded at the export gate. Observed prices will reflect the value of unobservable attributes valued by the buyer and/or the seller. For example, higher prices might reflect incentives paid to sellers to be reliable, while lower prices may arise when buyers provide a guaranteed demand. Researchers, and policy-makers, should be cautious about attributing, say, heterogeneous markups or incomplete pass-through of higher costs to undesirable forms of market power. Emran et al. (2021) illustrate this point. Prompted by concerns over abuse of market power, the government of Bangladesh banned "order traders," a certain type of intermediary in the edible oils market. However, because traders relax the credit constraints of wholesalers, the reform increased domestic prices and weakened the pass-through of imported crude prices.

The Prevalence and Value of Relationships

Many important aspects of trade exchanges are noncontractible and potentially subject to opportunistic behavior; indeed, some evidence from coffee markets suggests that half of observed defaults on forward contracts are caused by the exporter's reneging on promised deliveries to take advantage of improved market conditions (Blouin and Macchiavello 2019).[1] Even when a contract is in place, it is meant to clarify what parties expect from each other, with both sides knowing and expecting that the contract will not be enforced in court. Baker, Gibbons, and Murphy (2002) refer to these arrangements as relational contracts; that is, "informal agreements sustained by the value of future relationships." Under these circumstances, parties tend to stick with partners they trust. Long-term relationships based on trust have been documented in many settings, but weak institutions and limited contract enforcement might give them a particularly prominent role in developing and international markets (for a review, see Macchiavello 2022).

A significant share of international trade takes place in long-term relationships between buyers and sellers; indeed, the vast majority of US imports occur in pre-existing relationships (Monarch and Schmidt-Eisenlohr 2020; Monarch 2022). Using data from work in progress (Cajal-Grossi, Del Prete, and Macchiavello 2022; Del Prete et al. 2022), we construct a proxy for the prevalence of pre-existing relationships between exporters and foreign buyers in the global coffee and garment

[1] In many commodity markets there is a trade-off between insuring against price swings and counterparty risk. Parties can insure against market price risk by agreeing to a fixed price in advance. This however comes at the risk of one of the two parties reneging on the deal if spot market prices change sufficiently. Alternatively, parties can agree on price-indexed contracts that track market spot prices, foregoing insurance. Blouin and Macchiavello (2019) show that the possibility of defaults leaves many exporters of coffee uninsured against price risk.

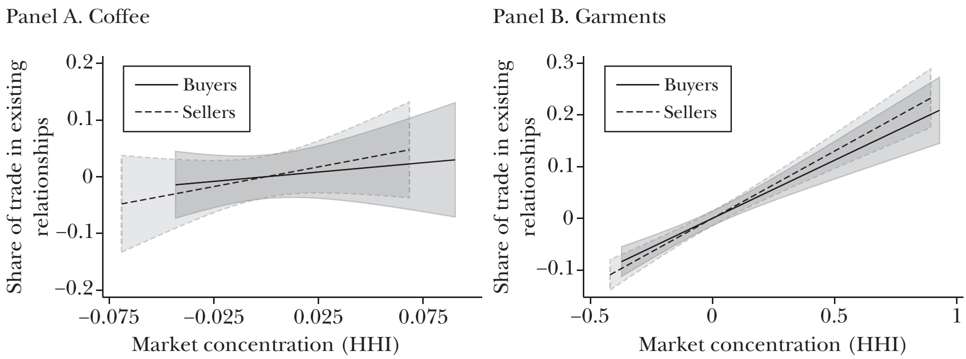
industries. In 2019, around 80 percent and 70 percent of trade in coffee and garments, respectively, took place between buyers and sellers that had traded the year before (see the online Appendix for details).[2]

The (future) value of the relationship deters parties from giving in to the temptation to behave opportunistically and deviate—yet this value is not directly observed. One approach to quantify the value of the relationship is to measure temptations to deviate. Macchiavello and Morjaria (2015) note that a relationship must be at least as valuable to the exporter as the extra revenues that the exporter could earn selling to a different available buyer at a higher price. They compute relationship values among Kenyan flower exporters using spot market prices at the Dutch auctions—a sales channel available to all—as a lower bound to the value of temptation. They find that the average long-term relationship with a foreign buyer in this market is worth about 30 percent of the exporter's yearly profits. Blouin and Macchiavello (2019) follow a similar approach to quantify the value of relationships in the coffee sector and find even larger estimates. A different approach is to measure profit margins earned from different buyers. This is difficult to do, as it requires observing both the prices earned from, and the costs incurred to supply, specific buyers. Cajal-Grossi, Macchiavello, and Noguera (forthcoming) relax these data constraints and find estimates of the value of relationships in the Bangladesh garment sector commensurate with those in Blouin and Macchiavello (2019) and Macchiavello and Morjaria (2015). These examples suggest that relationships in global supply chains can be valuable. Indeed, these estimates imply that, due to contracting problems, valuable trading opportunities do not take place because parties do not have sufficient "relationship value" to provide adequate incentives.

How do market power and relationships interact? A perfectly competitive market, without abnormal profits or rents, cannot sustain relationships. Figure 2 offers some suggestive evidence consistent with this hypothesis. The data underlying these figures correspond to distinct markets, defined in this case as product-origin combinations, where the product is an HS6 code and the origin is a country exporting coffee or garments. The horizontal axis reports the Herfindahl-Hirschman indices based on market shares of sellers and buyers—a proxy for concentration. The vertical axis reports the share of all exports in 2019 that occurred between parties that were observed trading in 2018, and thus have a pre-existing relationship. The left panel considers the case of coffee. As noted above, most green coffee is traded within a single HS6 code. Each data point in the figure corresponds to one of the 14 countries for which we have data. Despite the few observations, we find a positive, and statistically significant, correlation between market concentration and

[2] Vertical integration can also remedy the contracting problems discussed above and, indeed, accounts for a significant share of global trade (Antràs 2003). Vertical integration is almost entirely absent in garments. Large European and American retailers—even those that used to be garment manufacturers in their origin countries, such as Levy's and VF—own few factories abroad. In contrast, several international traders have integrated backward into exporting and processing stages of the coffee chain in sourcing countries (Del Prete et al. 2022).

Figure 2
Market Concentration and Relationships at the Export Gate



Source: Data are from Del Prete et al. (2022) and Cajal-Grossi, Del Prete, and Macchiavello (2022) respectively. See the online Appendix for further details.
Note: The figure shows the correlation between market concentration and the share of exports traded in relationships in coffee and garments. HHI stands for Herfindahl-Hirschman Index. Horizontal and vertical axes' variables are residualized against the size of the market, in terms of exported values. The linear fits over 14 observations in coffee and 1,113 observations in garments are presented alongside 95 percent confidence intervals.

the prevalence of relationships across coffee-sourcing origins that account for over 90 percent of global coffee trade.

The right panel considers the case of garments. Here, instead, we have data for seven origins that account for about one-third of developing countries' garments exports to the United States and Europe. We can however define markets more precisely, taking advantage of the numerous HS6 codes in garments. Again, the figure displays a positive and statistically significant correlation between market concentration and the prevalence of relationships.

Forming and Maintaining Relationships

Given that relationships are so widely used and appear to be valuable, natural questions arise: Where do these relationships come from? How are they sustained? And how do they influence market structure? While relationships can potentially bring benefits, they can also be used to sustain noncompetitive conduct like collusive arrangements (Bernasconi et al. 2023) or even to shut out potential entrants from markets.

Suitable partners for international trade are typically hard to find, and their discovery calls for costly efforts from both buyers and sellers (Eaton et al. 2022). Relational partners are not discovered by third-party reviews, exporters' directories, or attendance at industry meetings. Instead, in many cases, firms experiment with alternative trade partners until they settle on a relationship. The experimentation process can be uncertain, particularly in markets in which firms' operations are frequently disrupted by shocks. Studying garments in Bangladesh, Cajal-Grossi (2022) finds that buyers experiment to learn about potential suppliers. Following the Rana Plaza disaster in Bangladesh, international buyers concerned with potentially negative reputation spillover became more selective and reluctant to experiment.

Maintaining relationships also requires specific organizational capabilities. Multiple functions, ranging from design to distribution to human resources, must be coordinated across the entire organization to source inputs relationally from suppliers (Milgrom and Roberts 1990). Firms, even within narrowly defined industries, end up adopting very different approaches to sourcing (Helper and Henderson 2014). At one extreme, "spot" buyers spread purchases among multiple arm's-length suppliers, allocating short-term orders to the lowest bidders and bearing the costs of suppliers' nonperformance. At the other extreme, "relational" buyers allocate orders to a few suppliers with whom they develop long-term relationships (Taylor and Wiggins 1997).

Studying the garment sector in Bangladesh, Cajal-Grossi, Macchiavello, and Noguera (forthcoming) proxy for these sourcing strategies by exploiting the intuition that relational buyers concentrate sourcing among a relatively small number of suppliers. They obtain a cross-sectional characterization of buyers' sourcing strategies that maps closely to qualitative accounts in the industry. They find that a buyer's approach to sourcing is correlated across origins and products: buyer-level fixed effects explain a much larger share of the variation in sourcing strategies than the interaction of product with origin and destination markets fixed effects. This suggests that buyers' capabilities, rather than characteristics of the transactions (such as product complexity or the institutional quality of the sourcing country), are key determinants of sourcing practices.

Cajal-Grossi, Macchiavello, and Noguera (forthcoming) also show that a given exporter earns higher margins when supplying relational buyers as opposed to spot buyers. Using novel data that match quantities and prices of fabric and labor on sewing lines to specific export orders, they find that relational buyers pay higher prices for orders with similar product characteristics, including the quality, price and efficiency of the two inputs. In principle, relational buyers might thus be a vehicle for upgrading, enabling producers in developing countries to increase value addition through the provision of hard-to-contract upon attributes (such as reliability).

Beyond the Export Gate

Global supply chains reach down into the business relationships within domestic economies, as shown in Figure 1. We now turn to the domestic side of coffee and garment supply chains. Our emphasis is again on the importance of relationships: between smallholder coffee farmers and their domestic buyers and between garment factories and their workers. In both sectors, a relatively few large firms may command significant market power over farmers and workers. The market power of domestic processors and intermediaries in agricultural chains is often credited for low prices paid to farmers (Zavala 2022). Similarly, in many developing-country settings there are few large manufacturing firms offering industrial jobs (Hsieh and Olken 2014). Echoing Figure 2 above, we show examples in both chains suggesting that monopsony power is *positively* correlated with the quality of relationships.

Monopsony Power and Interlinked Transactions in Agriculture

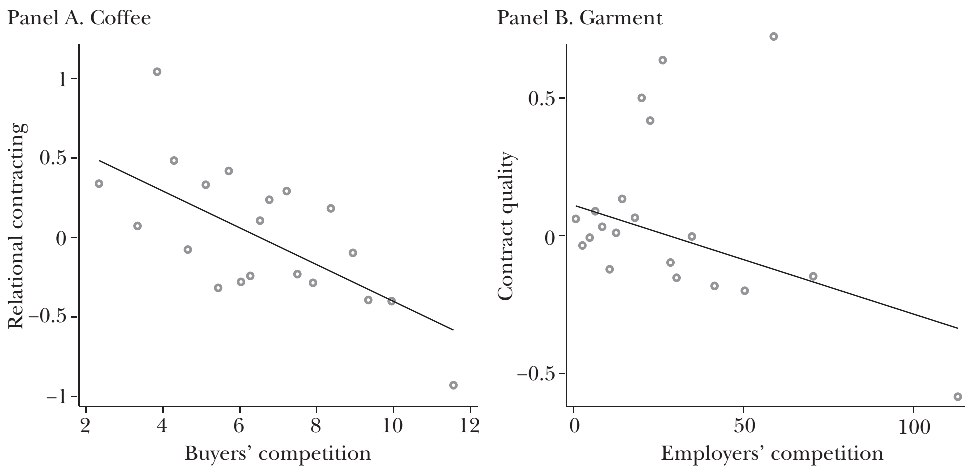
Farmers often face noncompetitive market structures downstream. For example, Bergquist and Dinerstein (2020) use an ingenious combination of experimental designs and structural modeling to study market conduct among agricultural traders in Kenya. Their estimates cannot reject collusive behavior—perhaps sustained by long-term relationships—among traders. In India, a law restricts farmers from selling their goods to intermediaries in their own state. Exploiting variation in competition induced by the law, Chatterjee (2023) shows that farmers are paid substantially lower prices when they face less competitive markets. In many cases—for example, tea, tobacco, sugar, and palm oil—smallholders' produce must be processed by firms that, due to fixed investments and high transport costs in rural areas, accrue substantial monopsonistic power over farmers. Rubens (2023) studies a policy reform that consolidated cigarette manufacturers in the Chinese tobacco industry; he finds that the reform increased manufacturers' market power over the farmers that sell tobacco leaf, distorting input markets, without generating significant gains in productivity.

Market power can also hinder quality upgrading. Using internal records from a large Colombian exporter, De Roux et al. (2022) document higher margins for the exporter on higher-quality coffee: while higher-quality coffee commands a significant price premium at the export gate, none of the price premium is passed on to domestic producers. An analysis of the pass-through of weather and exchange rate shocks to input and output prices reveals that the higher margin earned by the exporter on higher-quality coffee beans arises due to the exporter enjoying relatively higher market power in the upstream market for such beans relative to standard quality ones.

How do market power and relationships interact? Smallholder farmers in developing countries are likely to face imperfect domestic markets for inputs (Duflo, Kremer, Robinson 2008; Ashraf, Giné, and Karlan 2009; Duflo, Kremer, and Robinson 2011; Bold et al. 2017), credit (Karlan et al. 2014), insurance (Cai, Janvry, and Sadoulet 2015; Casaburi and Willis 2018), saving (Casaburi and Macchiavello 2019), and land (Acampora, Casaburi, and Willis 2022). In the presence of these market imperfections, farmers may enter "interlinked transactions" (Bardhan 1991) with their buyers, in which the sale of the produce is bundled with the provision of inputs and services. The underlying contracts are typically not enforceable in court, and so the interlinked transactions rely on long-term relationships.

For example, Macchiavello and Morjaria (2021) study the impact of competition between coffee-washing stations in the Rwanda coffee chain on the use of

Figure 3
Competition and Relationships in the Domestic Stage of the Supply Chain



Source: The data used in panel A are from Macchiavello and Morjaria (2021). Panel B's data are from Cajal-Grossi and Kreindler (2023).
Notes: Panel A shows the correlation between spatial competition and relational contracting between mills and farmers in the Rwanda coffee chain. The variables in both axes are residualized against a set of geographic controls. Panel B shows the correlation between spatial competition and relational contracting between garment factories and workers in the Bangladeshi garment chain. The relational contracting index is residualized against plant size (as measured by its own exports on the year of assessment). See online Appendix A for further details.

relational contracts. In this context, efficiency requires mills and smallholder farmers to exchange a bundle of services (inputs, credit, second payments, and others) before, during, and after the harvest season. The left-hand panel of Figure 3 uses coffee mills as the unit of observation. On the horizontal axis, the number of mills located within ten kilometers of a given mill offers a measure of the competition between buyers. The vertical axis shows an index of relational contracting between the mills and surrounding farmers. The index is calculated by combining information obtained from detailed surveys of both mills and random samples of farmers in the surrounding areas. The survey precisely measures the bundle of services that farmers and mills exchange before, during, and after the harvest season. Overall, competition in sourcing between mills is negatively correlated with the adoption of these relational practices, which is consistent with our earlier observation that rents are necessary to sustain valuable relationships. Macchiavello and Morjaria (2021) take advantage of an engineering model for the optimal placement of mills to create an instrumental variable for the level of competition and find that mills that face more competition use fewer relational contracts with farmers and exhibit worse performance. An additional competing mill also reduces the aggregate quantity of coffee supplied to mills by farmers and likely makes farmers worse off.[3]

In sum, there is the beginning of a consistent body of evidence suggesting that market power can lead to worse outcomes for farmers. But given that farmers also operate in a context with market imperfections, building and maintaining relationships between farmers and downstream players that have market power is crucial. We need to know more about how this is done. These buyers might need specific capabilities to develop such relationships in populations often characterized by a large number of smallholder farmers with low levels of education and general trust.

For example, farmer-owned cooperatives might take farmers' interests more into account and facilitate good relationships. On the other hand, cooperatives are fragile governance forms, due to redistributive pressures (Kremer 1997) and to capture (Banerjee et al. 2001). Despite their importance, we know relatively little about the functioning of cooperatives. Montero (2022) exploits a land reform in El Salvador that induced a discontinuous change in the probability of forming cooperatives. Relative to outside ownership (via haciendas), he finds that cooperatives perform relatively better in staple crops (such as maize and beans), whose output is not contractible (because farmers can easily hide output or consume it directly), than in cash crops (such as sugarcane and coffee), whose output can be more easily monitored—and thus redistributive pressures are more distortionary.

More broadly, a relational perspective emphasizes the difference between simply changing prices paid to farmers, which is relatively easy to do, as opposed to changing the equilibrium of the relationship with farmers, which is much harder. For example, Casaburi and Macchiavello (2015) document challenges in building clarity around the relational contract in a large dairy cooperative in Kenya. Abouaziza et al. (2023) find suggestive evidence that an intervention aimed at improving clarity around relational contracts in the Rwanda coffee chains increases loyalty, but only among the largest (and, arguably, more sophisticated) farmers.

Monopsony Power and Industrial Relations in Garments

The incorporation of developing countries into global value chains has increased productivity in manufacturing, and created better-paying jobs in these countries (World Bank 2020). In garments, these jobs have had broader societal benefits, especially for women, including delaying marriage and childbearing (Heath and Mobarak 2015), increasing female empowerment (Molina and Tanaka 2023), and improving health outcomes among children born to female workers (Atkin 2009). The effect of participation in global value chains on human capital accumulation is

[3] In the years following the survey upon which Macchiavello and Morjaria (2021) base their analysis, the industry kept witnessing the significant entry of new coffee mills and a further deterioration of relational contracts between mills and farmers. Many mills in the industry were acquired by downstream exporters. In follow-up work, Macchiavello and Morjaria (2022) show that this consolidation did not reduce prices earned by farmers but, when led by foreign-owned companies, led to improvements in the mill's efficiency and capacity utilization—possibly due to better management practices in building and managing relationships with farmers.

more ambiguous: the availability of these jobs can lead to reduced or to increased human capital formation depending on whether such jobs are low- or high-skill relative to alternatives (Heath and Mobarak 2015; Atkin 2016; Blanchard and Olney 2017; Li 2018). That said, while this conclusion is not well documented, it is plausible that participation in global value chains could support increased human capital formation in the longer-term.

Despite these potential benefits, especially in the garments sector, such jobs often entail very long work hours under difficult conditions. In a thought-provoking study, Blattman and Dercon (2018) randomized applicants to an industrial job offer in five large firms in Ethiopia. While the offer doubled initial exposure to industrial jobs, most workers quit within months. In fact, exposure to industrial job increased health problems. The high turnover rate potentially suggests that a set of complementary changes must occur for workers to benefit from this type of jobs. Indeed, the apparel sector is prone to conflict between firms and their workers, industrial disputes and labor unrest are frequent, and a high worker turnover is common. All this costs dearly to firms in terms of productivity. For example, in a rare case study of a Bangladeshi sweater factory that laid off 25 percent of its workforce following an episode of unrest, Akerlof et al. (2020) found a persistent productivity reduction (and income losses) among surviving workers, possibly due to a deliberate shading of performance to punish the factory's management.

Why do stronger and more stable relationships not emerge between firms and workers in the garment industry? One view is that poor industrial relations follow from firms' optimal responses to local conditions in the presence of an abundant supply of low-skill, homogeneous labor (Robinson 1962; Krugman 1997). An alternative view is that firms are operating inside the efficient production frontier and that building better relationships with workers might provide a win-win. As employment contracts are notoriously incomplete (Simon 1951), workers and employers must rely on relational contracts to sustain cooperation and improve performance (Baker, Gibbons, and Murphy 2002).

However, as already noted in our discussion about farmers, relational contracts are hard to build. This is so even within large, well-managed firms in industrialized countries (Gibbons and Henderson 2012). The parties need to develop trust and a clear understanding of each other's implicit commitments. The task is arguably harder for firms in developing countries, many of which have low productivity (Hsieh and Klenow 2009) and thus less margin of maneuver for building relational capabilities (Powell 2019). In turn, many key managerial practices rely on relational contracts between employers and employees. This may help explain why firms in developing countries adopt fewer management practices, as measured by the World Management Survey (Bloom et al. 2014), potentially further stifling productivity growth (Bloom et al. 2013). Globally, apparel firms lag behind other manufacturing firms in their adoption of management practices (authors' calculations using the World Management Survey data).

This line of thought suggests that supporting firms in developing countries to build better relationships with workers might be a win-win: boosting firm performance while upgrading workers' job quality. In this vein, there is a small but growing research agenda on improving industrial relations in the garment sector. Adhvaryu, Kala, and Nyshadham (2023) experimentally test the returns to investing in workers' soft skills—leadership, communication, teamwork, and collaboration—in the context of a large, Indian firm. They find productivity gains of 13.5 percent among trained workers, positive spillovers to peers consistent with increased cooperation, and a 256 percent net return of the program to the firm eight months after completion.

Improving information flows between workers and managers and increasing workers' voice inside the firm can also support better industrial relations. On the former, Boudreau et al. (2023) investigate secure survey methods designed to monitor harassment in organizations. Under standard direct-reporting systems, workers will hesitate to report harassment for fear of retaliation. They conduct a survey experiment with workers employed by a large Bangladeshi firm. They find that providing plausible deniability of such reports through "hard garbling," or randomly flipping some "no" responses to "yes," has large effects on reporting of harassment. Adhvaryu, Molina, and Nyshadham (2022) examine the value of giving voice to workers in a large Indian apparel firm. After what proved to be a disappointing minimum wage hike, they invited randomly selected workers to provide feedback on their job conditions, supervisor's performance, and job satisfaction. Enabling voice in this manner reduced turnover and absenteeism, showing that workers inherently value voice at work.

In the above discussion of Adhvaryu, Kala, and Nyshadham (2023) finding benefits from teaching soft skills, the alert reader may have noticed that there was no mention of workers' wages—which in fact did not increase despite the substantial productivity gains to the employer from the intervention. This finding is consistent with employers in developing countries having labor market power (for discussion, see Amodio and de Roux 2021). Besides the small numbers of large employers, high search frictions (Abebe, Caria, and Ortiz-Ospina 2021), limited information about employers' quality (Boudreau, Heath, and McCormick forthcoming), and limited workers' mobility (Méndez and Van Patten 2022) all contribute to employers' labor market power. From a policy perspective, quantifying the relative importance of these forces appears crucial. Studying the Bangladesh garment sector, Cajal-Grossi and Kreindler (2023) use high-frequency surveys and a spatial model of workers' job location decisions, to show large welfare losses from distance-driven information frictions.

Gender norms can exacerbate the negative consequences of employers' monopsony power on female workers and thus limit the garment sector's potential to foster women's empowerment. Sharma (2023) argues that the impact of employers' monopsony power varies across gender and that this can account for a sizeable share of the gender wage gap in the textile and apparel industry in Brazil. Similarly, Menzel and Woodruff (2021) find that a significant share of the gender pay gap in Bangladeshi factories is due to women's lower external mobility and internal promotion rates. On the latter, although women account for over 90 percent of the workers in the sewing section of large garment factories in Bangladesh, they account for only 5 percent of the line supervisors and lower-level managers (Macchiavello et al. 2020). A randomized controlled trial that encouraged factories to promote more women to line supervisory roles reveals that inaccurate beliefs—possibly inherited from the early days of the industry in which women had not yet entered the labor force—are partly responsible for the underpromotion of women in the industry.

Again, we can ask how market power and "relational contracts" interact. Leveraging data from Cajal-Grossi and Kreindler (2023), the right-hand panel of Figure 3 (presented earlier) looks at employer competition and employer–employee relationships in 290 garment plants in urban Bangladesh. The horizontal axis measures the extent of employer competition for workers, based on the number of active exporters within one kilometer of the plant. The vertical axis shows an index of the quality of the relational contract between the plant and its workers. The index encompasses a large number of nonpecuniary job attributes that are valuable to workers but that plants may renege upon, such as the advance notice for overtime, conflict resolution mechanisms, and the compliance with dismissal protocols, all obtained from audits conducted by the Better Work program of the International Labour Organization (see details in the online Appendix). As in the case of the coffee supply chain, there is a negative correlation between employers' competition for workers and the quality of relational contracting between workers and employers.[4]

Relationships and Sustainable Supply-Chains

Based on case studies in coffee and garments, we focus on how a relational approach at the export gate can improve the quality of relationships in the domestic portion of the chain—for example, paying higher prices to farmers in coffee and ensuring safer working conditions in garments. We believe that these insights are likely to be relevant for other industries as well as to broader issues, including environmental conservation and preventing sourcing that fuels armed conflict. Insofar as the state has limited capacity to regulate and monitor social and environmental standards in developing countries, the role of buyers may be especially important. Although the evidence on this topic is still emerging, we believe it offers a particularly valuable direction for future research.

Concurrent with the rise of global value chains, numerous nonprofit organizations have established certification programs to deal with sustainability challenges.

[4] Along similar lines, Méndez and Van Patten (2022) provides a fascinating study of the long–term impact of a large land concession held by the United Fruit Company—a large monopsonist in Costa Rica—from 1899 to 1984. Using a geographic regression discontinuity design, they document how the United Fruit Company had a positive, persistent effect on living standards due to its investment in local amenities for the workforce. They also show that where workers were more mobile, the local investment effort United Fruit Company was higher.

Table 1
Selected Initiatives in Coffee and Garments

Initiative	*Type*	*Sectoral and Geographic Scope*	*Targets*
Panel A. Coffee in Colombia			
Fair Trade https://www.fairtrade.net/ Start: 1997 Status: (Active)	Fairtrade International: Multi-stakeholder non-profit association FLOCERT: Private limited certification company	Agriculture (25 countries)	Better prices, working conditions, terms of trade, local sustainability
Rainforest Alliance https://www.rainforest-alliance.org/ Start: 1987 Status: (Active)	International nonprofit	Business, Agriculture, Forests (70 countries)	Climate, human rights, livelihoods, forests
The Common Code for the Coffee Community (4C) https://www.4c-services.org/ Start: 2003 Status: (Active)	Multi-stakeholder initiative (producers, industry, civil society organizations)	Coffee (20 countries)	Economic, social, and environmental sustainability
Nespresso AAA Sustainable Quality™ Program https://www.sustainability.nespresso.com/aaa-sustainable-quality-program Start: 1997 Status: (Active)	Program run by private corporation	Coffee (18 countries)	Quality, productivity, social and environmental sustainability
Panel B. Garments in Bangladesh			
Accord on Fire and Building Safety in Bangladesh https://bangladeshaccord.org Start: 2013 Status: (Active from outside the country)	Multi-stakeholder initiative (brands, retailers, labor unions)	Apparel, tertiary sectors (Bangladesh)	Health and safety
Alliance for Bangladesh Worker Safety http://bangladeshworkersafety.org/ Start: 2013 Status: (Not active)	Multi-firm initiative	Apparel, tertiary sectors (Bangladesh)	Health and safety
Nirapon https://www.nirapon.org/ Start: 2019 Status: (Active from outside the country)	Multi-firm initiative	Apparel, tertiary sectors (Bangladesh)	Health and safety
Action on Living Wages (ACT) https://actonlivingwages.com/ Start: 2017 Status: (Active)	Multi-stakeholder initiative (brands, retailers, labor unions)	Apparel, textile, footware (4 countries)	Wages, freedom of association, purchasing practices
Better Work https://betterwork.org/ Start: 2014 Status: (Active)	Public-private partnership (ILO-IFC, brands, plants)	Apparel, tertiary sectors (10 countries)	Social compliance (all)
Fair Labor Association https://www.fairlabor.org/ Start: 1999 Status: (Active)	Multi-stakeholder initiative (brands, retailers, universities, suppliers, civil society organizations)	Manufacturing, agriculture (nonspecific)	Labor standards (all)

Source: All information comes from the publicly available, official webpages of the initiatives (included beneath the Initiative's name, in column 1).
Note: The table presents the authors' systematization of a number characteristics of selected multi–stakeholder initiatives addressing different sustainability dimensions in the coffee (panel A) and garments (panel B) supply chains.

Table 1 provides information on selected initiatives applicable to the coffee and garments value chains. Fairtrade International, the first example in panel A, launched its Fairtrade Certification Mark in 2002 (for a review in this journal, see Dragusanu, Giovannucci, and Nunn's article in the Summer 2014 issue). The Fairtrade system has two key aspects: a price premium paid to a farmer organizer or producer; and for a few products, including coffee, a minimum price guaranteed when products are sold as Fairtrade. In both cases, the ensuing price premium must be spent for "social projects" in the community. The second example, the Rainforest Alliance, is an international nongovernmental organization focused on forest preservation and the livelihoods of farmers and forest communities. It certifies agricultural and forestry products, as well as tourism businesses, based on environmental, social, and livelihood-based criteria.

Despite the growth of these certification schemes, there is relatively little rigorous evidence about their effects. In coffee, Dragusanu, Montero, and Nunn (2022) find gains for producers and farmholders in Costa Rica, but not for unskilled workers. Other studies are less optimistic. For example, De Janvry, McIntosh, and Sadoulet (2015) show that farmers pay to have all their produce certified, but only a share of their produce is sold as such, and so price premia over the entire production are limited. In reviewing the literature, Oya, Schaefer, and Skalidou (2018) argue that the evidence is mixed and the impact likely context-specific, and that better evaluation designs are needed to understand the impact of these schemes.

Turning to buyers, many have developed their own labor and environmental standards for upstream suppliers, using methods including codes of conduct for suppliers, buyer-driven certification programs, and industry initiatives, among others. But as already discussed, aspects such as social and environmental compliance are notoriously hard to monitor for the buyer—and thus difficult to include in formal contracts enforceable in courts in a cost-effective manner. In practice, this means that buyers adopting relational sourcing strategies at the export gate may be better placed to enforce social and environmental standards in their supply chains and/or to succeed in achieving upgrading in these areas when they attempt to enforce it. While more evidence is needed to confirm this hypothesis, the examples of the AAA Sustainable Quality Program in coffee and the example of the retailer Gap in garments point in that direction.

The flagship buyer-driven program in the coffee industry is the Nespresso AAA Sustainable Quality Program. The firm, a multinational buyer, combines contractual arrangements at the export gate with training and agricultural extension services to farmers, to ensure that it can reliably purchase large volumes of high-quality coffee from farmers. Notably, the contract with the exporter specifies the price (premium) that must be paid to the upstream farmers. This arrangement, a form of vertical restraint, counterbalances the monopsonistic power of the exporter, which, as shown in De Roux et al. (2022), would tend to set price premia for high-quality beans too low. Looking at this program in Colombia's coffee industry, Macchiavello and Miquel-Florensa (2019) find that it leads to significant investment and quality upgrading in the chain. The program increased the total surplus by about 30 percent, with at least half of the gains going to farmers. The vertical restraint aspect of the program plays a crucial role in fostering quality upgrading.

In many export-oriented agricultural supply chains, long-run sustainability considerations as well as regulatory and activist threats have made buyers particularly concerned with the environmental impact of their supply chain operations.[5] For example, Henderson and Nellemann (2011) describe Unilever's pivot toward environmentally sustainable sourcing, its motivations, and the implementation challenges it faced. Palm oil and cocoa—among others—are important drivers of deforestation in Africa and South Asia (Balboni et al. forthcoming). Ensuring the environmental sustainability of supply chains poses even harder challenges than issues of higher prices or pay. Unlike with low prices and poor working conditions—for which one could imagine that the farmer, or the worker, has the incentive to report violations of an agreement—the structure of incentives is less obvious in cases where the worker or community benefits, at least in the short-term, from environmental degradation, or in which they may simply lack information about whether or how agreed-upon rules are being violated. Coordinated trade policy has been proposed as a method to curb deforestation, but there are doubts about monitoring and enforcement of such rules (Dominguez-Iino 2021; Hsiao 2022).[6]

On the labor side, some evidence suggests that exporting can lead to better working conditions; for example, see Tanaka (2020) for an analysis in the Myanmar garment sector. Buyers' approaches to improving working conditions vary, but they often include minimum labor standards coupled with monitoring by means of compliance audits, developing remediation plans for violations, and monitoring of remediation. For example, Amengual and Distelhorst (2019) conduct a case study of Gap, a multinational apparel retailer that primarily sources from suppliers in developing countries and that maintains its own supplier code of conduct for labor and environmental issues, which is enforced by its supplier responsibility department. The authors study a change in Gap's policies that more strongly conditioned its trade with suppliers on their labor audit performance. Prior to the change, there was no effect of a failing audit grade on suppliers' future compliance, while afterward, a failing grade led suppliers to improve compliance by 0.8 standard deviations. In the relational contract, providing incentives to suppliers to adopt better labor standards requires conditioning trade on cooperation with these standards.

Even when such initiatives have the intended positive effect on workers in the buyer's supply chain, their overall impact is more nuanced. Alfaro-Urena et al. (2022) develop a general equilibrium model to study the incidence of foreign buyers'

[5] In practice, many brands have found that the willingness of consumers to pay for environmentally sustainable products to be quite low, except in specific cases—such as the garment brand Patagonia—that have successfully targeted niche markets.

[6] In coffee, quality upgrading and environmental sustainability go hand-in-hand. The Nespresso AAA Sustainable Quality Program was developed in partnership, and shares environmental standards, with Rainforest Alliance. Macchiavello and Miquel-Florensa (2019) point to other research documenting the environmental benefits of practices similar to those in the program.

responsible sourcing policies and show that the welfare implications are a priori ambiguous, due to the interaction of a terms-of-trade effect and input market distortions. They estimate the model in the context of Costa Rica and find that responsible sourcing significantly increased the welfare of the 21 percent of low-wage workers employed at exposed suppliers, but at the cost of real income losses of –2.2 percent to the remaining 79 percent of low-wage workers. Alfaro-Urena et al. (2022) make a valuable step forward, but evidence on the industry- and economy-level impacts of buyer sustainability interventions remains very thin, and more evidence is certainly needed in this area.

Increasingly, social and environmental standards in global value chains are being set and monitored by multi-buyer and multi-stakeholder initiatives. In the multi-buyers approach, there is an agreement to commit to common standards along their supply chains. In the case of multi-stakeholder initiatives, nongovernment organizations, labor unions, or other nonprofit-oriented organizations participate in setting standards and monitoring their implementation. Both models have certain advantages over single-buyer approaches; for example, they improve administrative efficiency by harmonizing standards, curbing free-riding problems, and enabling monitoring of participating buyers by civil society (in the case of multi-stakeholder initiatives). They may also be subject to certain drawbacks, such as concerns about lowest-common-denominator standards and facilitating coordination over pricing. Examples of multi-stakeholder initiatives that cover parts of the garments global value chain include the Fair Labor Association and the International Labor Organization's Better Work Program; see panel B of Table 1 for more information.

Two prominent examples in the garments chain that also appear in Table 1 are the Accord and the Alliance occupational safety and health initiatives in Bangladesh, which were established in response to the Rana Plaza collapse in 2013 (mentioned in the introduction). At its peak, the Accord included over 200 primarily European apparel buyers and labor unions. The Alliance was an initiative of 29 primarily North American apparel buyers. Together, these initiatives covered most of Bangladesh's apparel sector. Between the collapse and July 2016, the International Labor Organization (2017) reported that 3,780 factories were inspected for safety; of these, 59 percent were audited (and subsequently monitored on their remediation) by the Accord or the Alliance, which were estimated to cover 75 percent of the sector in terms of its direct employment. An important feature of both initiatives was that they provided incentives for suppliers to adopt stronger occupational safety and health standards through buyers' unilateral termination of sourcing relationships with suppliers that failed to cooperate.

In addition to building safety, the buyer initiatives enforced a local mandate for occupational safety and health committees that was passed in the aftermath of the collapse. Boudreau (2021) randomized the roll-out of the Alliance's enforcement intervention for the mandate across 84 supplier factories. She documents that the Alliance's intervention increased suppliers' compliance with the mandate. Exploiting experimental variation in the strength of occupational safety and health committees, she shows that they improved workers' health and safety. These improvements did not come at a cost to workers in terms of wages or employment, nor to factories in terms of labor productivity; indeed, the estimated effects on labor productivity are positive. The results are consistent with implementation of occupational safety and health committees not being very costly either for employers or with employers exercising labor market power. Interestingly, the effects are stronger for factories that had better management practices at baseline, which is consistent with the earlier argument that low capabilities may constrain firms in developing countries from building stronger relationships with workers.

Action on Living Wages is an agreement between 19 multinational buyers and a global union that aims to ensure living wages (that is, the minimum income required for workers to meet their basic needs) in the textile and apparel value chains. The agreement aims to achieve this goal through collective bargaining at the industry level, freedom of association, and responsible sourcing practices. One distinguishing feature of this initiative is its focus on buyers' purchasing practices that affect workers' wages and working conditions. Participating buyers commit to work toward itemizing labor costs in their purchase orders with suppliers in a way that adheres to the initiative's costing protocols. This type of arrangement thus echoes the relational vertical restraint implemented by the AAA Nespresso Program studied in Macchiavello and Miquel-Florensa (2019).

An important concern with the provision of enforcement capacity by nongovernmental actors is that it may crowd out provision by the state, which may be counterproductive to the development of state capacity in the longer term. In the context of the occupational safety and health initiatives in Bangladesh, there was coordination between the buyer initiatives and the International Labour Organization, with the latter supporting the government to build its capacity by focusing on the share of the sector that fell outside the purview of the buyer initiatives. While this type of coordination seems desirable, its effectiveness remains an open question. A related concern is that nongovernmental enforcement initiatives may even threaten the sovereignty of the states in which they operate. In Bangladesh, the Accord's authority to operate was eventually challenged in court by a domestic firm that had one of its factories unilaterally terminated from supplying to Accord members; in a protracted court battle, the Accord fought to operate in Bangladesh. Ultimately, Bangladesh's High Court ruled that the Accord had to vacate the country, although it continues to operate from abroad. This example illustrates the types of political economy concerns that can arise when powerful downstream buyers participate in enforcing labor standards in developing countries.

Conclusion

We have explored the economics of two prominent value chains: coffee and garments. We discussed several aspects of exchange between exporters and buyers in these value chains that are not accounted for in standard international trade datasets, which describe quantities and prices of goods as they cross national borders; consequently, these sources provide incomplete accounts of the functioning of global value chains, both at the import and export gates, as well as in the domestic parts of the chain. Leveraging contextual knowledge and originally collected data is needed to overcome these limitations.

We have emphasized the realities of incomplete contracts and imperfect markets. Well-functioning relationships may be able to increase both efficiency and equity for the participation of developing countries in global value chains. But these relationships are hard to establish and sustain. Relationships also alter how markets function: to understand market power along supply chains, frameworks should adequately account for the underlying contractual frictions that relationships address. We strongly suspect that these themes are relevant to many other areas of international trade beyond the two industries of focus in this paper.

We have also argued that relational approaches at the export gate can be leveraged to improve the efficiency and equity of supply chains in domestic markets and potentially contribute to addressing a variety of urgent sustainability challenges. In both coffee and apparel, large foreign buyers that source relationally at the export gate have shown some ability to improve prices to farmers, and workers' conditions among suppliers. Policymakers in destination countries are showing a growing interest in initiatives that aim at regulating environmental and social standards in supply chains, including certain provisions of the Wall Street Reform and Consumer Protection Act of 2010 (often known as the Dodd-Frank act) on conflict minerals in the United States, the German Supply Chain Act, and the European Union Directive. These initiatives highlight the importance of grounding the design of policies aimed at promoting sustainable sourcing practices in a deeper, evidence-driven understanding of the impact of, constraints to, and form of participation in global value chains in developing countries. At present, however, we have very limited rigorous evidence about the impact of sustainability standards driven by buyers and nongovernmental organizations; beyond directly impacted suppliers, workers, and farmers, more evidence is needed on their industry and economy-wide effects and on their longer-run implications for developing countries.

In addition, a much deeper understanding of why buyers opt for different strategies, and how fixed these decisions are, can inform policy in other ways. There seems to be a lot of money left on the table in the form of unrealized gains from trade due to contracting problems. From the perspective of developing countries, attracting foreign buyers that will invest in relationships along their supply chains may be a promising direction for trade and development policy. More broadly, a relational approach might also foster the resilience of supply chains. In garments, relational buyers are less diversified across sourcing origins (Cajal-Grossi, Del Prete, and Macchiavello 2022); a relational approach might thus be a substitute strategy for diversification to foster supply chain resilience.

■ *We are grateful to editors, Heidi Williams, Erik Hurst, Nina Pavcnik, and Timothy Taylor, for their insightful feedback and their help in improving this piece. We thank Richard Baldwin and Bob Gibbons for many useful suggestions, and Davide del Prete for sharing data. Aishwarya Rebelly and Canyon Keanu Can provided outstanding research assistance. Financial support from ERC Grant 818767 Sharing Gains is gratefully acknowledged.*

References

Abebe, Girum, A. Stefano Caria, and Esteban Ortiz-Ospina. 2021. "The Selection of Talent: Experimental and Structural Evidence from Ethiopia." *American Economic Review* 111 (6): 1757–1806.

Abouaziza, Mohamed, Rocco Macchiavello, Iris Steenkamp, and Ameek Singh. 2023. "Building Relational Contracts in the Rwanda Coffee Sector." Unpublished.

Acampora, Michelle, Lorenzo Casaburi, and Jack Willis. 2022. "Land Rental Markets: Experimental Evidence from Kenya." NBER Working Paper 30495.

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. 2023. "Returns to On-the-Job Soft Skills Training." *Journal of Political Economy*. https://doi.org/10.1086/724320.

Adhvaryu, Achyuta, Teresa Molina, and Anant Nyshadham. 2022. "Expectations, Wage Hikes and Worker Voice." *Economic Journal* 132 (645): 1978–93.

Akamatsu, Kaname. 1962. "A Historical Pattern of Economic Growth in Developing Countries." *Developing Economies* 1 (S1): 3–25.

Akerlof, Robert, Anik Ashraf, Rocco Macchiavello, and Atonu Rabbani. 2020. "Layoffs and Productivity at a Bangladeshi Sweater Factory." CESifo Working Paper 8492.

Alfaro-Urena, Alonso, Benjamin Faber, Cecile Gaubert, Isabela Manelici, and Jose P. Vasquez. 2022. "Responsible Sourcing? Theory and Evidence from Costa Rica." NBER Working Paper 30683.

Amengual, Matthew, and Greg Distelhorst. 2019. "Can Sourcing Help Enforce Global Labor Standards? Evidence from the Gap Inc Supply Chain." Unpublished.

Amodio, Francesco, and Nicolás de Roux. 2021. "Labor Market Power in Developing Countries: Evidence from Colombian Plants." IZA Discussion Paper 14390.

Antràs, Pol. 2003. "Firms, Contracts, and Trade Structure." *Quarterly Journal of Economics* 118 (4): 1375–1418.

Antràs, Pol. 2020. "Conceptual Aspects of Global Value Chains." *World Bank Economic Review* 34 (3): 551–74.

Antràs, Pol, and C. Fritz Foley. 2015. "Poultry in Motion: A Study of International Trade Finance Practices." *Journal of Political Economy* 123 (4): 853–901.

Ashraf, Nava, Xavier Giné, and Dean Karlan. 2009. "Finding Missing Markets (and a Disturbing Epilogue): Evidence from an Export Crop Adoption and Marketing Intervention in Kenya." *American Journal of Agricultural Economics* 91 (4): 973–90.

Atkin, David. 2009. "Working for the Future: Female Factory Work and Child Health in Mexico." Unpublished.

Atkin, David. 2016. "Endogenous Skill Acquisition and Export Manufacturing in Mexico." *American Economic Review* 106 (8): 2046–85.

Bai, Jie, Ludovica Gazze, and Yukun Wang. 2021. "Collective Reputation in Trade: Evidence from the Chinese Dairy Industry." *Review of Economics and Statistics* 104 (6): 1121–37.

Baker, George, Robert Gibbons, and Kevin J. Murphy. 2002. "Relational Contracts and the Theory of the Firm." *Quarterly Journal of Economics* 117 (1): 39–84.

Balboni, Clare, Aaron Berman, Robin Burgess, and Benjamin A. Olken. Forthcoming. "The Economics of Tropical Deforestation." *Annual Review of Economics*.

Baldwin, Richard E., and Philippe Martin. 1999. "Two Waves of Globalisation: Superficial Similarities, Fundamental Differences." NBER Working Paper 6904.

Banerjee, Abhijit V., and Esther Duflo. 2014. "Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program." *Review of Economic Studies* 81 (2): 572–607.

Banerjee, Abhijit, and Kaivan Munshi. 2004. "How Efficiently Is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur." *Review of Economic Studies* 71 (1): 19–42.

Banerjee, Abhijit, Dilip Mookherjee, Kaivan Munshi, and Debraj Ray. 2001. "Inequality, Control Rights, and Rent Seeking: Sugar Cooperatives in Maharashtra." *Journal of Political Economy* 109 (1): 138–90.

Bardhan, Pranab, ed. 1991. *The Economic Theory of Agrarian Institutions*. Oxford, UK: Oxford University Press.

Bergquist, Lauren Falcao, and Michael Dinerstein. 2020. "Competition and Entry in Agricultural Markets: Experimental Evidence from Kenya." *American Economic Review* 110 (12): 3705–47.

Bernasconi, Mario, Miguel Espinosa, Rocco Macchiavello, and Carlos Suarez. 2023. "Relational Collusion in the Colombian Electricity Market." CESifo Working Paper 10384.

Blanchard, Emily J., and William W. Olney. 2017. "Globalization and Human Capital Investment: Export Composition Drives Educational Attainment." *Journal of International Economics* 106: 165–83.

Blattman, Christopher, and Stefan Dercon. 2018. "The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia." *American Economic Journal: Applied Economics* 10 (3): 1–38.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *Quarterly Journal of Economics* 128 (1): 1–51.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, Daniella Scur, and John Van Reenen. 2014. "JEEA–FBBCA Lecture 2013: The New Empirical Economics of Management." *Journal of the European Economic Association* 12 (4): 835–76.

Bloom, Nicholas, and John Van Reenen. 2010. "Why Do Management Practices Differ across Firms and Countries?" *Journal of Economic Perspectives* 24 (1): 203–24.

Blouin, Arthur, and Rocco Macchiavello. 2019. "Strategic Default in the International Coffee Market." *Quarterly Journal of Economics* 134 (2): 895–951.

Bold, Tessa, Kayuki C. Kaizzi, Jakob Svensson, and David Yanagizawa-Drott. 2017. "Lemon Technologies and Adoption: Measurement, Theory and Evidence from Agricultural Markets in Uganda." *Quarterly Journal of Economics* 132 (3): 1055–1100.

Boudreau, Laura. 2021. "Multinational Enforcement of Labor Law: Experimental Evidence on Strengthening Occupational Safety and Health (OSH) Committees." Unpublished.

Boudreau, Laura E., Sylvain Chassang, Ada Gonzalez-Torres, and Rachel M. Heath. 2023. "Monitoring Harassment in Organizations." NBER Working Paper 31011.

Boudreau, Laura, Julia Cajal-Grossi, and Rocco Macchiavello. 2023. "Replication data for: Global Value Chains in Developing Countries: A Relational Perspective from Coffee and Garments." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E191904V1.

Boudreau, Laura, Rachel Heath, and Tyler McCormick. Forthcoming. "Migrants, Experience, and Working Conditions in Bangladeshi Garment Factories." *Journal of Economic Behavior and Organization*.

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet. 2015. "Social Networks and the Decision to Insure." *American Economic Journal: Applied Economics* 7 (2): 81–108.

Cajal-Grossi, Julia. 2022. "Searching for Trade Partners in Developing Countries." Unpublished.

Cajal-Grossi, Julia, Davide Del Prete, and Rocco Macchiavello. 2022. "Sourcing Strategies and Supply-Chain Resilience in the Global Garments Industry." Unpublished.

Cajal-Grossi, Julia and Gabriel Kreindler. 2023. "Workers in Space: Evidence from Urban Bangladesh." Unpublished.

Cajal-Grossi, Julia, Rocco Macchiavello, and Guillermo Noguera. Forthcoming. "Buyers' Sourcing Strategies and Suppliers' Markups in Bangladeshi Garments." *Quarterly Journal of Economics*.

Casaburi, Lorenzo, and Rocco Macchiavello. 2015. "Loyalty, Exit, and Enforcement: Evidence from a Kenya Dairy Cooperative." *American Economic Review* 105 (5): 286–90.

Casaburi, Lorenzo, and Rocco Macchiavello. 2019. "Demand and Supply of Infrequent Payments as a Commitment Device: Evidence from Kenya." *American Economic Review* 109 (2): 523–55.

Casaburi, Lorenzo, and Jack Willis. 2018. "Time versus State in Insurance: Experimental Evidence from Contract Farming in Kenya." *American Economic Review* 108 (12): 3778–3813.

Chatterjee, Shoumitro. 2023. "Market Power and Spatial Competition in Rural India." *Quarterly Journal* *of Economics*. https://doi.org/10.1093/qje/qjad004.

de Janvry, Alain, Craig McIntosh, and Elisabeth Sadoulet. 2015. "Fair Trade and Free Entry: Can a Disequilibrium Market Serve as a Development Tool?" *Review of Economics and Statistics* 97 (3): 567–73.

de Roux, Nicolás, Rocco Macchiavello, Pepita Miquel-Florensa, and Eric Verhoogen. 2022. "Markups and Quality: Evidence from Colombian Coffee Exports." Unpublished.

Del Prete, Davide, Pablo Fajgelbaum, Amit Khandelwal, and Rocco Macchiavello. 2022. "It's Not Just Coffee: Market Power in the Global Value Chains." Unpublished.

Dominguez-Iino, Tomás. 2021. "Efficiency and Redistribution in Environmental Policy: An Equilibrium Analysis of Agricultural Supply Chains." Unpublished.

Dragusanu, Raluca, Eduardo Montero, and Nathan Nunn. 2022. "The Effects of Fair Trade Certification: Evidence from Coffee Producers in Costa Rica." *Journal of the European Economic Association* 20 (4): 1743–90.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2008. "How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya." *American Economic Review* 98 (2): 482–88.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101 (6): 2350–90.

Eaton, Jonathan, David Jinkins, James R. Tybout, and Daniel Xu. 2022. "Two-Sided Search in International Markets." NBER Working Paper 29684.

Emran, M. Shahe, Dilip Mookherjee, Forhad Shilpi, and M. Helal Uddin. 2021. "Credit Rationing and Pass-Through in Supply Chains: Theory and Evidence from Bangladesh." *American Economic Journal: Applied Economics* 13 (3): 202–36.

Gereffi, Gary. 1999. "International Trade and Industrial Upgrading in the Apparel Commodity Chain." *Journal of International Economics* 48 (1): 37–70.

Gereffi, Gary, John Humphrey, Raphael Kaplinsky, and Timothy J. Sturgeon. 2001. "Introduction: Globalisation, Value Chains and Development." *IDS Bulletin* 32 (3): 1–8.

Gibbons, Robert, and Rebecca Henderson. 2012. "Relational Contracts and Organizational Capabilities." *Organization Science* 23 (5): 1350–64.

Gresser, Charis, and Sophia Tickell. 2002. *Mugged: Poverty in Your Coffee Cup*. Oxford: Oxfam International.

Heath, Rachel, and A. Mushfiq Mobarak. 2015. "Manufacturing Growth and the Lives of Bangladeshi Women." *Journal of Development Economics* 115: 1–15.

Helper, Susan, and Rebecca Henderson. 2014. "Management Practices, Relational Contracts, and the Decline of General Motors." *Journal of Economic Perspectives* 28 (1): 49–72.

Henderson, Rebecca M., and Frederik Nellemann. 2011. "Sustainable Tea at Unilever." Harvard Business School Marketing Unit Case 712-438.

Hsiao, Allan. 2022. "Coordination and Commitment in International Climate Action: Evidence from Palm Oil." Unpublished.

Hsieh, Chang-Tai, and Benjamin A. Olken. 2014. "The Missing 'Missing Middle.'" *Journal of Economic Perspectives* 28 (3): 89–108.

Hsieh, Chang-Tai, and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics* 124 (4): 1403–48.

International Coffee Council (ICO). 2020. *Coffee Development Report 2020*. London: International Coffee Organization.

International Labor Organization. 2017. *Strengthening Worplace Safety and Labour Rights in the Bangladesh Ready-Made Garment Sector.* Geneva: International Labor Organization.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry. 2014. "Agricultural Decisions after Relaxing Credit and Risk Constraints." *Quarterly Journal of Economics* 129 (2): 597–652.

Kremer, Michael. 1997. "Why Are Worker Cooperatives So Rare?" NBER Working Paper 6118.

Krugman, Paul. 1997. "In Praise of Cheap Labor." *Slate*, March 21. https://slate.com/business/1997/03/in-praise-of-cheap-labor.html.

Ksoll, Christopher, Rocco Macchiavello, and Ameet Morjaria. 2022. "Electoral Violence and Supply Chain Disruptions in Kenya's Floriculture Industry." *Review of Economics and Statistics*. https://doi.org/10.1162/rest_a_01185.

Li, Bingjing. 2018. "Export Expansion, Skill Acquisition and Industry Specialization: Evidence from China." *Journal of International Economics* 114: 346–61.

Locke, Richard M. 2013. *The Promise and Limits of Private Power: Promoting Labor Standards in a Global*

Economy. Cambridge, UK: Cambridge University Press.

Macchiavello, Rocco. 2022. "Relational Contracts and Development." *Annual Review of Economics* 14: 337–62.

Macchiavello, Rocco, Andreas Menzel, Atonu Rabbani, and Christopher Woodruff. 2020. "Challenges of Change: An Experiment Promoting Women to Managerial Roles in the Bangladeshi Garment Sector." NBER Working Paper 27606.

Macchiavello, Rocco, and Josepa Miquel-Florensa. 2017. "Vertical Integration and Relational Contracts: Evidence from the Costa Rica Coffee Chain." CEPR Discussion Paper 11874.

Macchiavello, Rocco, and Josepa Miquel-Florensa. 2019. "Buyer-Driven Upgrading in GVCs: The Sustainable Quality Program in Colombia." CEPR Discussion Paper 13935.

Macchiavello, Rocco, and Ameet Morjaria. 2015. "The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports." *American Economic Review* 105 (9): 2911–45.

Macchiavello, Rocco, and Ameet Morjaria. 2021. "Competition and Relational Contracts in the Rwanda Coffee Chain." *Quarterly Journal of Economics* 136 (2): 1089–1143.

Macchiavello, Rocco, and Ameet Morjaria. 2022. "Acquisitions, Management, and Efficiency in Rwanda's Coffee Industry." NBER Working Paper 30230.

Manova, Kalina, and Zhihong Yu. 2016. "How Firms Export: Processing vs. Ordinary Trade with Financial Frictions." *Journal of International Economics* 100: 120–37.

Méndez, Esteban, and Diana Van Patten. 2022. "Multinationals, Monopsony, and Local Development: Evidence from the United Fruit Company." *Econometrica* 90 (6): 2685–2721.

Menzel, Andreas, and Christopher Woodruff. 2021. "Gender Wage Gaps and Worker Mobility: Evidence from the Garment Sector in Bangladesh." *Labour Economics* 71: 102000.

Milgrom, Paul, and John Roberts. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *American Economic Review* 80 (3): 511–28.

Molina, Teresa, and Mari Tanaka. 2023. "Globalization and Female Empowerment: Evidence from Myanmar." *Economic Development and Cultural Change* 71 (2): 519–65.

Monarch, Ryan. 2022. "'It's Not You, It's Me': Prices, Quality, and Switching in U.S.-China Trade Relationships." *Review of Economics and Statistics* 104 (5): 909–28.

Monarch, Ryan, and Tim Schmidt-Eisenlohr. 2020. "Longevity and the Value of Trade Relationships." Unpublished.

Montero, Eduardo. 2022. "Cooperative Property Rights and Development: Evidence from Land Reform in El Salvador." *Journal of Political Economy* 130 (1): 48–93.

Oya, Carlos, Florian Schaefer, and Dafni Skalidou. 2018. "The Effectiveness of Agricultural Certification in Developing Countries: A Systematic Review." *World Development* 112: 282–312.

Powell, Michael. 2019. "Productivity and Credibility in Industry Equilibrium." *RAND Journal of Economics* 50 (1): 121–46.

Robinson, Joan. 1962. *Economic Philosophy*. Piscataway, NJ: Transaction Publishers.

Rubens, Michael. 2023. "Market Structure, Oligopsony Power, and Productivity." Unpublished.

Sharma, Garima. 2023. "Monopsony and Gender." Unpublished.

Simon, Herbert A. 1951. "A Formal Theory of the Employment Relationship." *Econometrica* 19 (3): 293–305.

Tanaka, Mari. 2020. "Exporting Sweatshops? Evidence from Myanmar." *Review of Economics and Statistics* 102 (3): 442–56.

Taylor, Curtis R., and Steven N. Wiggins. 1997. "Competition or Compensation: Supplier Incentives under the American and Japanese Subcontracting Systems." *American Economic Review* 87 (4): 598–618.

Verhoogen, Eric. Forthcoming. "Firm-Level Upgrading in Developing Countries." *Journal of Economic Literature.*

Watkins, Kevin, and Penny Fowler. 2002. *Rigged Rules and Double Standards: Trade, Globalisation, and the Fight against Poverty.* Oxford: Oxfam International.

World Bank. 2020. *World Development Report 2020: Trading for Development in the Age of Global Value Chains.* Washington, DC: World Bank.

Zavala, Lucas. 2022. "Unfair Trade? Monopsony Power in Agricultural Value Chains." Unpublished.

Journal of Economic Perspectives—Volume 37, Number 3—Summer 2023—Pages 87–110

Are Developed Countries Outsourcing Pollution?

Arik Levinson

Since the beginning of serious national environmental regulatory policy in the early 1970s, public debate has assumed that, absent constraints, polluting industries would shift production to countries with less strict standards. Developed countries would outsource pollution by importing goods they once produced at home. Domestic industries in developed countries have argued against the strongest environmental standards, for fear of losing jobs and market share to countries with weaker standards. Environmental advocacy groups have worried that outsourcing will worsen pollution in developing countries; the environmental policy jargon for that is "pollution havens." At the same, proponents of both protecting the environment and allowing unrestricted international trade suspect that fears of outsourcing are being used to justify both weaker pollution standards and protectionist trade barriers.

At the 1972 Stockholm Conference on the Human Environment, developed countries asked the United Nations to set uniform environmental rules to help them avoid losing their polluting industries to pollution havens. Since then, we have seen periodic proposals to take unilateral action to prevent outsourcing pollution from US firms to overseas suppliers. Walter Cronkite (1980), the iconic long-time anchor of the CBS Evening News, advocated banning imports from countries with less strict environmental rules to "protect both American industry and the environment."

In 1991, Oklahoma Senator David Boren proposed the International Pollution Deterrence Act. That bill would have imposed a tax, or tariff, on imports of goods

■ *Arik Levinson is Professor of Economics, Georgetown University, Washington, DC. His email address is arik.levinson@georgetown.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.3.87.

manufactured in countries without strict standards. Similar worries arose among US states. The 1977 amendments to the Clean Air Act were in part an attempt to prevent outsourcing by states with strict environmental regulations to states with less strict standards (Pashigian 1985).

By 1993, when Canada, the United States, and Mexico were negotiating details of the North American Free Trade Agreement (NAFTA), opposition by prominent environmental groups obliged the parties to include an environmental side agreement, the North American Agreement on Environmental Cooperation. A goal of that 1993 pact was to prevent industries from avoiding their own countries' environmental rules by outsourcing. Since then, international trade negotiations typically feature debates about the environmental consequences of lowering tariffs. Today the World Trade Organization has a standing committee with a broad mandate to address links between trade and the environment.

For pollutants that damage local environments, like hazardous waste or airborne particulates, outsourcing by developed countries would involve tradeoffs. Environmental quality would improve in the developed world and decline in the pollution havens. When the focus of concern turns to climate change, outsourcing raises a different issue. Carbon dioxide and other greenhouse gases cause the same environmental damage no matter where in the world they are emitted. Thus, if national policies to reduce carbon emissions only lead the sources of those emissions to relocate to other countries, with no effect on total global carbon pollution, that outsourcing would undermine the environmental benefits of those national policies (Campbell, McDarris, and Pizer 2021). The climate-policy jargon for that is "leakage."

In July 2021, with the goal of combating leakage, US legislators introduced a bill to impose a tax, or tariff, on imports.[1] The tariff rate would be based on the domestic cost that US climate regulations impose on US production—although, ironically, at the time the United States had no comprehensive federal climate policies, at least not ones that imposed easily calculable costs. Today, the Inflation Reduction Act of 2022 mostly grants subsidies to US firms for reducing greenhouse gases. Subsidies raise no concerns about leakage, because companies have no incentive to relocate overseas to avoid them. Despite that fact, some US environmental organizations like the Sierra Club (2022) support versions of the idea—taxing foreign goods based on their carbon content without regard to US domestic carbon policy.

Europe does have rules that impose costs on carbon pollution through its Emissions Trading System. As of March 2023, permits were selling for more than €100 per ton of carbon emitted. To address concerns about leakage from that system, in April 2023 the European Parliament approved key features of a new "carbon border adjustment mechanism." It will levy tariffs, starting at low rates in 2026, that will eventually equal the cost of permits businesses would have needed to purchase had they

[1] For an overview of the FAIR Transition and Competition Act of 2021, see https://www.coons.senate.gov/imo/media/doc/one_pager__fair_transition_and_competition_act_-_117.pdf (accessed May 31, 2022).

manufactured the imported goods in Europe. The goal is to eventually eliminate the ability of European businesses to avoid Europe's climate policy by outsourcing.

To date, however, no rules against outsourcing pollution have taken effect; indeed, laws that impose tariffs to offset less strict environmental standards in other countries may be illegal under existing international trade rules (Böhringer et al. 2022). But notice that all such efforts presume that deterrence is necessary to prevent polluting industries from evading developed countries' strict environmental standards—whether aimed at local or global concerns—by relocating production to places were pollution is regulated less strictly and exporting products back to the developed countries. In this essay, I demonstrate that evidence for that presumption is sparse.

The next two sections set the stage for more detailed empirical discussion. I first point out some general patterns: real GDP in high income countries nearly doubled in the last three decades, but local and global pollutants from those countries have not risen. Can this be explained by the rapid rise in trade between high-income countries and the rest of the world? I then explore the theoretical question that hovers in the background: from an economic perspective, what's wrong with outsourcing pollution? As we will see, the answer depends in part on the nature of the pollutant.

The following sections then address the title question by splitting it into three separate parts. First, have high-income countries improved their own environments by importing goods produced in polluting factories? Second, has pollution worsened in countries manufacturing those goods for export to high-income countries? Third, have the environmental regulations enacted by rich developed countries had a cause-and-effect relationship on either of the first two changes?

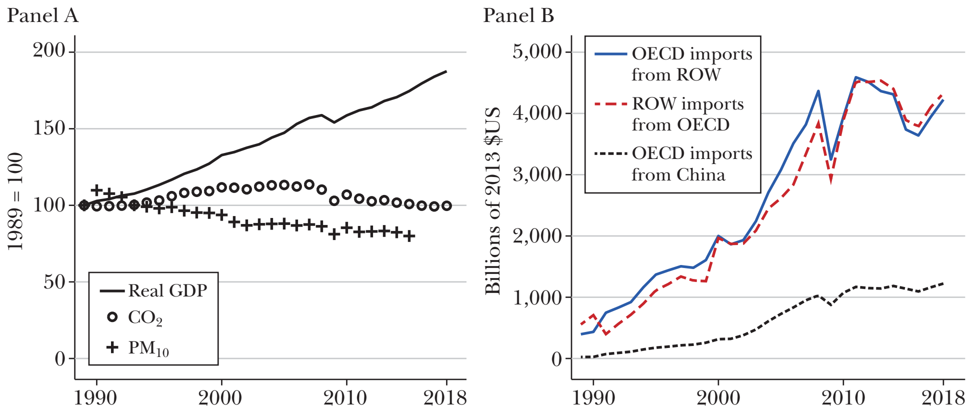
Trends in Growth, Pollution, and Trade

In recent decades, high-income countries as a group have managed to grow their economies without emitting more pollution. In theory, outsourcing could explain that achievement. In practice, basic patterns of trade and pollution cast doubt on that explanation.

As a starting point, consider the group of 24 high-income countries that belonged to the Organisation for Economic Co-operation and Development (OECD) as of 1993 (before the organization expanded to 38 members, including some middle-income countries).[2] Figure 1a plots the real gross domestic product (GDP) of those 24 countries, indexed so that 1989 equals 100. Their collective

[2] The 24 members of the OECD as of 1993 were Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Japan, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Türkiye, United Kingdom, and the United States. The 14 countries that have joined since then are Chile, Colombia, Costa Rica, Czech Republic, Estonia, Hungary, Israel, Korea, Latvia, Lithuania, Mexico, Poland, Slovak Republic, and Slovenia. List is available at https://www.oecd.org/about/document/ratification-oecd-convention.htm.

Figure 1
OECD Cleanup and Trade



Source: In panel A GDP in $US from the Organisation for Economic Co-operation and Development (OECD) (data.oecd.org) is adjusted using the US GDP deflator from the Federal Reserve (fred.stlouisfed.org). Pollution of particles smaller than 10 microns (PM_{10}) and carbon dioxide (CO_2) are taken from the Emissions Database for Global Atmospheric Research (edgar.jrc.ec.europa.eu). Panel B depicts trade volumes from the United Nations (comtrade.un.org), as reported by the importing countries.
Note: Real GDP doubled while pollution remained flat or declined (panel A). Trade between OECD countries and the rest of the world increased more than seven-fold (panel B).

economic output nearly doubled in the past three decades, which you can see by following the top solid line from its starting point at 100 in 1989 to nearly 200 by 2018. For comparison, Figure 1a also includes two measures of air pollution. Particulate matter smaller than 10 microns (PM_{10}) can be viewed as representing local air pollution, and is represented by small crosses in Figure 1a. Carbon dioxide (CO_2) dioxide is a global pollutant, shown as small circles. Both PM_{10} and CO_2 remained nearly flat or even declined in those 24 countries, even as their collective GDP doubled.

Could imports to the 24 OECD countries from the rest of the world help to explain the pattern of growth without additional pollution? Figure 1b plots the value of goods imported and exported by those 24 countries. The solid line plots total imports for these 24 high-income countries from the rest of the world (ROW), measured in real US dollars. Those imports grew from around $400 billion in 1989 to more than $4 trillion by 2018. The dashed line plots trade in the reverse direction, from the rest of the world (ROW) to these 24 countries.

As Figure 1b shows, trade between those 24 OECD countries and the rest of the world is eight to ten times larger today than it was 30 years ago—a growth rate much faster than the doubling of overall economic activity. In addition, notice that goods trade between the 24 countries and the rest of the world was approximately balanced. The solid line depicting OECD imports from the rest of the world closely follows the dashed line depicting the reverse. (Remember, the group of 24 high-income counties includes both economies that have often had large trade deficits, like the United States, but also countries that have had large trade surpluses, like Japan and Germany.) Thus, if outsourcing explains pollution reductions in developed countries, it must be that the types of goods the OECD imports from the rest of the world differ from the types of goods other countries import from the OECD.

Finally, Figure 1b illustrates some nuance in the definition of "developed" because of the way economies and trade patterns have changed in the last few decades. For example, China, which accounts for a large and growing fraction of imports by the 24 high-income countries, rapidly transitioned from "low-income" to "upper middle-income" over this period, according to the World Bank. It seems plausible that countries like China may be changing the types of goods they export and import as they develop.

How have the 24 high-income OECD countries managed to increase the scale of their real output without increasing local or global pollution, as shown in Figure 1a? There are two possibilities here: technology and composition. Technology—sometimes dubbed "technique"—would involve the countries producing largely the same mix of goods, but using production processes that pollute less: cleaner fuels, more energy-efficient equipment, or better end-of-pipe controls. Composition would involve the high-income countries producing a different mix of goods and services, shifting towards products requiring less pollution to manufacture. In turn, that composition change itself could be traced to two causes. Either citizens of high-income countries consume a cleaner mix of products, or they purchase the ones that create the most pollution from importers rather than from domestic manufacturers—that is, they outsource pollution.

To put these explanations in more specific terms, if a country manufactures more steel and cars in factories that pollute less, that is technique. If a country's economy shifts from producing steel and cars to producing fewer polluting goods or services—electronics or insurance—that is composition. A shift in composition could result from the country's residents demanding less steel and fewer cars and more electronics and insurance. Or, that composition shift could result from the country importing the steel and cars it once produced domestically. Only the last explanation for the pattern in Figure 1a, importing polluting goods formerly produced domestically, involves outsourcing.

The next section will consider the economic argument for what can be wrong with outsourcing pollution. We then explore an accounting exercise to investigate how the much of the pattern in Figure 1a might be explained by outsourcing.

What's Wrong with Outsourcing Pollution?

Before describing the problem with outsourcing pollution, it's worth being clear about what is meant by each word: "outsourcing" and "pollution." Economists sometimes describe outsourcing as occurring at the firm level, when a particular company contracts with a third party to purchase goods or services—either a final output or intermediate inputs—that were previously produced or could have been produced by the company's own employees. A university that hires a landscaping company rather than employing groundskeepers has outsourced those jobs. In this essay, I use "outsourcing" in the more general national sense, to describe what happens when a country imports goods rather than producing them domestically. Those could be final products such as automobiles, or intermediate inputs to those products such as car parts, or inputs to those intermediate inputs such as steel and rubber.

As for pollution, in the popular conception it is an output, because it comes "out" of smokestacks or wastewater pipes. In economic terms, however, pollution is an input. Manufacturing a product for sale often requires pollution, just as it requires capital and labor. If it helps, think of pollution as a waste disposal service. Goods can be manufactured using more pollution and less capital and labor, or less pollution and more of those other inputs. Manufacturers can outsource pollution to countries with different production methods or lower standards just as they can outsource labor to countries with lower wages, by importing the goods created using that pollution and labor.

The problem with outsourcing pollution depends on the nature of the pollutant. If the pollution is local, like the emissions of particulates depicted by PM_{10} in Figure 1a, outsourcing improves the environment in one country and degrades it in another. If the pollution is global, like CO_2, shifting production from one country to another has no effect on the environment unless the countries' production methods involve different amounts of pollution. Before the world's attention turned to CO_2 emissions and climate change, most environmental policies targeted local pollutants: urban smog or airborne particulates, toxic waste dumps, and water pollution.

Outsourcing local pollution from one location to another may be undesirable, but it is not *necessarily* so. Just as some countries have comparative advantages from natural resource abundance or skilled labor, others may have comparative advantages in production of goods that cause pollution. Imagine an island country where pollution blows out to sea and rapidly disperses, doing little or no harm to its residents or anybody else. Trade between other countries and the island could expand production possibilities in both places.

Of course, most countries are not windswept islands, and most local pollution does at least some local damage. Even so, some outsourcing of local pollution might still be efficient, at least in the cold calculus of economics. If clean air is a normal good, meaning higher-income people want more of it, then citizens of rich democracies will vote for more stringent pollution policies than citizens of poor democracies. As a result, poor countries will have a comparative advantage in production that pollutes—and could become pollution havens. The World Bank's chief economist once made that efficiency argument in defense of pollution havens, to foreseeable controversy (a story told in Hausman and McPherson 2006, p. 12).

However, a long catalog of assumptions must be met before this efficiency result holds in full. Oates and Schwab (1996) formalized the theoretical case for outsourcing to pollution havens, in an academic setting that attracted less public scrutiny. They described countries competing to attract businesses that cause local pollution but raise local wages. The workers who suffer from the pollution must also be receiving the higher wages. Local governments must be welfare-maximizing, and their residents homogenous. Any industry profits need to be earned by those homogenous worker-residents, not paid to multinationals in other countries. If any of those conditions are not met, local regulatory authorities will set economically inefficient pollution standards. Those could be insufficiently strict, race-to-the-bottom, pollution-haven standards. Or they could be overly stringent, race-to-the-top, not-in-my-backyard (NIMBY) standards (Levinson 2003).

If some countries do not have sufficiently capable regulatory infrastructures necessary to manage their local environments, the odds that outsourcing pollution could lead to efficiency gains look even slimmer. If regulators fail to internalize externalities appropriately, through incompetence, corruption, or everyday politics, then importing goods from those places least capable of enacting and enforcing reasonable pollution regulations seems more likely to exacerbate market failures than to be Pareto-improving.

If we add concerns about equity and democratic representation, problems with outsourcing local pollution compound. Not everyone can vote at a ballot box or vote with their feet by emigrating if they find their preferences unrepresented, their wages insufficient, or the nearby pollution excessive.

With these concerns about local environments in mind—efficiency, regulatory capability, and equity—116 countries met in Switzerland in 1989 to adopt the Basel Convention on the Control of Transboundary Movements of Hazardous Wastes.[3] By shipping waste to developing countries, especially in Africa, developed countries had been outsourcing one pollution-intensive part of their manufacturing, the final disposal of dangerous byproducts. The Basel Convention recognized the problems with the "limited capabilities of the developing countries to manage hazardous waste." It explicitly prohibited the export of such waste from OECD to non-OECD countries.

Hazardous waste is a local pollutant for which the outsourcing process is particularly obvious and identifiable, because the pollution travels by ship between countries. For most pollutants, firms in high-income countries can simply invest in production processes that pollute in other countries or purchase their products directly. Local environmental quality would improve in high-income countries and degrade elsewhere. It would be good to know how much, if any, outsourcing of local pollution is happening as a result of those less obvious processes.

For global pollution like CO_2 that causes climate change, it does not matter whether the factory sits on a windswept island or in the center of a densely populated valley. If an industry that emits CO_2 relocates from a developed country to a developing one, without changing its production processes, environmental quality

[3] For background on the Basel Convention, see www.basel.int (accessed March 25, 2022).

does not improve or worsen anywhere. Regulations designed to combat climate change, but that merely shift pollution among countries, impose costs with no benefits. So understanding the degree to which developed countries outsource CO_2 is central to understanding the degree to which their domestic climate policies have had any success.

For local pollutants, outsourcing to pollution havens raises concerns about efficiency, regulatory capability, and equity. For global pollutants, in addition to those issues, outsourcing takes the form of leakage that would undermine the efficacy of domestic policies.

Are Developing Countries Cleaning Up by Outsourcing Pollution?

Assessing the extent to which high-income countries reduce domestic pollution by importing goods whose production generates foreign pollution is essentially an accounting exercise. Calculate how much total pollution is used to manufacture each product in developed countries. Divide that total pollution produced by the total dollar value of each product manufactured to get each product's pollution intensity, measured in tons of pollution per dollar of product sold. Multiply those pollution intensity values by the total value of imports for each good to get the pollution displaced by those imports. That tells us how much pollution each of those imported goods would have caused in the developed country had they been produced domestically instead of imported. Sum those multiples across all imported goods to get the total amount of pollution embodied in imports. That is the amount of pollution "outsourced."

For our purposes, we want to compare the amount of pollution outsourced by high-income countries to those with lower incomes with the amount of pollution outsourced (by this same definition) from lower-income to high-income countries. This exercise requires no identification of cause and effect. It is just a descriptive, multi-step accounting of trade flows.

Emission Intensities for Each Industry

The first step in that accounting requires a measure of how much pollution each industry emits. For example, the National Emissions Inventory (NEI) from the US Environmental Protection Agency reports US emissions of many dozens of air pollutants, including local pollutants like PM_{10} and global pollutants like CO_2, across nearly 300 different manufacturing industries.[4]

That level of detail is important to studying outsourcing of pollution. Consider the paper subsector, defined by the North American Industrial Classification System

[4]For a description, see Yang et al. (2017). For details, see www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei (accessed 4/8/2022).

(NAICS) with the three-digit code 322.[5] That includes pulp mills that convert raw wood into paper, NAICS six-digit industry code 322110, which is one of the more polluting processes in all of manufacturing. But it also includes factories that purchase paper and use it to manufacture envelopes and other stationery, NAICS 322230, a process that involves relatively little pollution. Knowing only that a developed country imports paper, without knowing if the shipment contains envelopes or the raw paper from pulp mills, is not sufficient to assess the pollution content or the degree of outsourcing.

Most studies of pollution outsourcing focus on the manufacturing sector. The other sectors are either not as polluting—finance, retail trade, warehousing—or they cannot easily be outsourced—construction, transport, electric utilities. In 2017, the North American Industrial Classification System listed 21 three-digit manufacturing subsectors comprising 360 six-digit manufacturing industries. Table 1 describes some sample industries from the National Emissions Inventory, ordered by their PM_{10} emissions intensities in column 1. Manufacturing stationery involves relatively little pollution. The pulp mills that make the paper involve a lot more.

Some countries aside from the United States do publish emissions inventories; however, they are typically less detailed than the US National Emissions Inventory. For example, the European Union reports emissions of some of those same air pollutants, including PM_{10} and CO_2, for about 75 industrial sectors, including manufacturing, mining, agriculture, and utilities.[6] Canada publishes a detailed emissions inventory for CO_2, for 142 industrial sectors, 87 of which are in manufacturing. But the US National Emissions Inventory covers the most pollutants and with the most narrowly detailed definitions of manufacturing industries.

Direct and Total Emissions Intensities

Next, for each good imported into a developed country, in order to know the amount of outsourced pollution we need to know the pollution that would have been emitted domestically while manufacturing that imported product, as well as the pollution that would have been emitted manufacturing inputs to that product, plus the pollution emitted manufacturing inputs to those inputs, and so on. Suppose that a developed country switches from producing envelopes domestically using domestically manufactured paper and begins importing envelopes that use paper manufactured elsewhere. If we use only the *direct* emissions intensities in column 1 of Table 1 to calculate outsourced pollution, and include only the pollution used to manufacture the envelopes themselves, that calculation would miss pollution that would have been

[5] For an overview of NAICS, see www.census.gov/naics/ (accessed 4/12/2002). NAICS codes are six digits long. The first two digits describe broad sectors, such as agriculture or manufacturing. The first three digits define subsectors, like textiles or transportation equipment within manufacturing. The four-digit codes are industry groups, like shipbuilding or motor vehicles within transportation. The full six-digit codes describe specific industries, such as those manufacturing truck trailers or brake systems.

[6] For discussion of the EU reporting, see www.eea.europa.eu/data-and-maps/data/industrial-reporting-under-the-industrial-4 (accessed 4/12/2022).

Table 1
Sample Emissions Intensities

NAICS	Industry	1,000 kg per 2013 $US of production: PM_{10} Direct	PM_{10} Total	CO_2 Direct	CO_2 Total	Percent of US manufacturing output 2021
334516	Analytical Laboratory Instrument Manufacturing	0.000006	0.08	3.4	198	0.41
334112	Computer Storage Device Manufacturing	0.000044	0.05	2.2	147	0.10
339910	Jewelry and Silverware Manufacturing	0.000078	0.37	6.2	340	0.14
336112	Light Truck and Utility Vehicle Manufacturing	0.000086	0.16	9.4	342	1.95
322230	Stationery	0.000210	0.21	13.3	507	0.06
333415	Air-Conditioning, Heating, and Refrigeration Equipment	0.000346	0.17	9.4	303	0.70
326160	Plastics Bottle Manufacturing	0.000637	0.29	13.1	682	0.22
335911	Storage Battery Manufacturing	0.001127	0.42	41.3	565	0.14
336414	Guided Missile and Space Vehicle Manufacturing	0.001631	0.04	2.1	97	0.28
311513	Cheese Manufacturing	0.003952	2.24	26.9	516	0.58
337127	Institutional Furniture Manufacturing	0.009099	0.20	18.4	369	0.09
311111	Dog and Cat Food Manufacturing	0.009550	2.22	42.1	493	0.43
325510	Paint and Coating Manufacturing	0.003914	0.39	98.0	592	0.53
331420	Copper Rolling, Drawing, Extruding, and Alloying	0.006126	0.73	36.0	505	0.39
322110	Pulp Mills	0.262743	0.68	328.9	827	0.13
327310	Cement Manufacturing	1.473377	1.82	7,199.0	8,077	0.20

Source: Emissions intensities from the EPA Environmentally-Extended IO Tables, www.epa.gov/land-research/us-environmentally-extended-input-output-useeio-technical-content (Yang et al. 2017). Percentages of US manufacturing output from the 2021 Annual Survey of Manufactures https://www2.census.gov/programs-surveys/asm/data/2021/ (file AM1831BASICC01).
Note: Just a few of the 200+ manufacturing industries in the EPA Environmentally-Extended IO Tables. Units are industrial emissions in 1,000 kilograms per 2013 US dollars. Direct numbers include only emissions from manufacturing final products. Totals include all upstream emissions from manufacturing inputs.

emitted by domestic pulp mills to manufacture the paper used as inputs to those envelopes, pollution emitted by the inputs to the pulp mills, and so on.

That sounds like a daunting problem. But in the 1970s, long before any of this data was even available, Leontief (1970) explained how to do the calculation. It requires an input-output table of the economy—that is, a table in which each row represents an output and each column represents an input. Each cell of the table reports how many dollars' worth of the corresponding input industry is used to manufacture one dollar's worth of the output industry. To cover all 360 six-digit manufacturing NAICS codes, an input-output table needs 360 × 360 = 129,600 entries. Leontief provided an example that also includes pollution as one of the inputs. He used only a few broadly defined sectors, all that was empirically possible at the time, but demonstrated the linear algebra techniques needed for a solution.

These days, the US Commerce Department publishes input-output tables for hundreds of industries that can be matched with six-digit NAICS codes. Combining those with the emissions intensities from the National Emissions Inventory, using Leontief's (1970) computational approach, yields a set of *total* industry-specific emissions intensities, which report the pollution needed to create a dollar's worth of each good, including all of the inputs, inputs to inputs, and so on. I walk through the details in Levinson (2009). Thankfully, the National Emissions Inventory now does that calculation for us and reports both direct and total emissions intensities.

Table 1 lists, for that same subset of six-digit NAICS industries, the total pollution from making the final product and all its inputs. For the stationery industry, the total PM_{10} intensity is 1,000 times larger than the direct intensity. For pulp mills that make the paper, total pollution is only 2.5 times the direct pollution. That pattern makes sense, because the highly polluting pulp mills are an important input into stationery, but not vice versa.

Linking Emissions Intensities to Exports and Imports by Industry

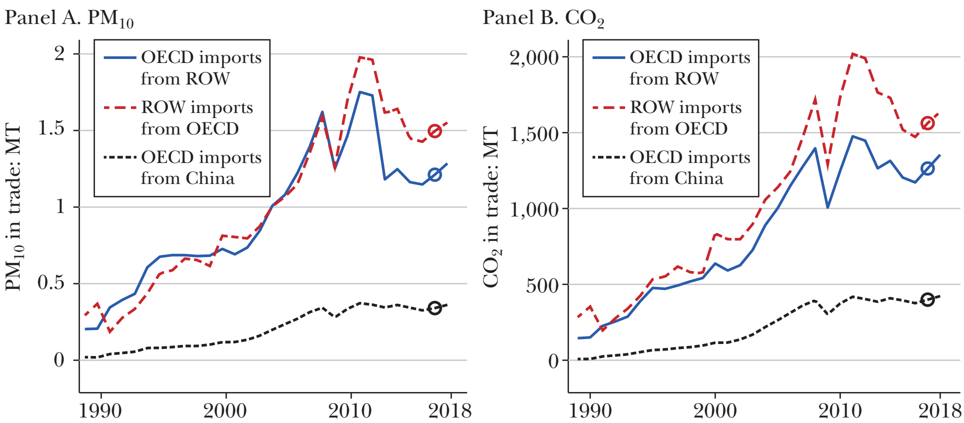
To estimate pollution outsourcing, the next requirement is data on each industry's imports from and exports to each country. The UN Comtrade data reports annual trade between every pair of countries, valued in US dollars, for thousands of commodities.[7] Those dollar values formed the basis for Figure 1b. I match those Comtrade commodity codes to the NAICS industries used by the emissions inventory using a concordance constructed by Pierce and Schott (2012). I then adjust those to 2013 dollars to match the emissions intensities reported in the National Emissions Inventory, using the US GDP deflator.

From there, the last step multiplies the total emissions intensity for each industry by the total dollar value of imports. That product provides the total amount of pollution outsourced when that particular industry's goods are imported. Summing those products across all imported industries yields the total amount of outsourced pollution embodied in those imports.

Figure 2 depicts pollution outsourcing as calculated by this method for two air pollutants: PM_{10} and CO_2. The solid blue lines estimate the pollution that would have been caused in the 24 high-income OECD countries to manufacture the goods that these countries actually imported from the rest of the world (abbreviated ROW).

[7] For details on the UN Comtrade data, see comtrade.un.org (accessed 12/17/2021).

Figure 2
Total Pollution in Trade



Source: Trade data from comtrade.un.org for 24 high-income countries that were in the OECD as of 1973 and the rest of the world (ROW). Pollution calculated using total emissions coefficients for the year 2017 (circled) from www.epa.gov/land-research/us-environmentally-extended-input-output-useeio-technical-content, using the method described in the text.
Note: MT=million tons.

Importantly, recognize that changes over time depicted in Figure 2 do not involve changes to production or pollution control technologies, because the data points for all years are calculated using the 2017 pollution intensities reported by the US EPA. The figure only shows changes in the total volume of imports and exports—the scale of trade—and shifts in which industries make up that total—its composition. The actual pollution that would have been caused in the OECD would differ if those pollution intensities—techniques—changed over time or differed in other countries.

The scale and composition estimates of pollution embodied in OECD imports, depicted in Figure 2, grew steeply over the past 30 years, from 0.2 to 1.29 million tons of PM_{10}, and from 145 to 1,354 million tons of CO_2. Wealthy countries now import goods responsible for six to nine times as much pollution as 30 years ago.

Evaluating the Outsourcing Claim: Absolute versus Percentages, Net versus Gross

On the surface, that result seems like straightforward evidence of outsourcing. But the rest of the world also imports goods from the high-income countries, and those imports have increased as well. The dashed red lines estimate the pollution actually emitted in OECD countries to manufacture the goods shipped in the reverse direction, exported from the OECD to the rest of the world. Using the same estimation approach, the dashed red lines in Figure 2 show that pollution caused by manufacturing goods in the OECD for export to the rest of the world has increased

Table 2
Pollution in Trade: 1989–2018

	OECD imports from ROW	ROW imports from OECD	OECD imports from China
Imports (2013 US $billion)			
1989	396	556	24
2018	4,225	4,331	1,223
Percent change	+966%	+678%	+4,961%
PM_{10} (Million tons)			
1989	0.20	0.29	0.02
2018	1.29	1.55	0.36
Percent change	+531%	+429%	+1,721%
CO_2 (Million tons)			
1989	145	284	8
2018	1,354	1,637	423
Percent change	+830%	+477%	+5,117%

Note: This table reports the starting and ending values, for 1989 and 2018, of the lines in Figure 2.

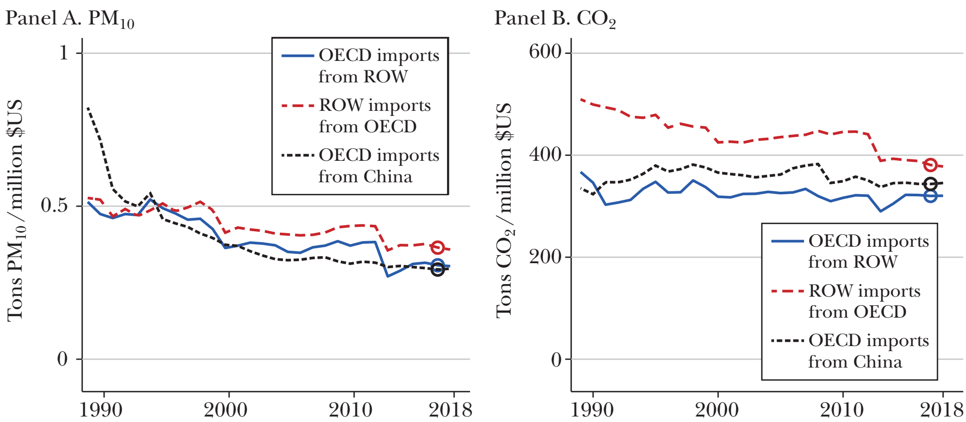
from 0.29 to 1.55 million tons of PM_{10}, and from 284 to 1,637 million tons of CO_2. Wealthy countries now *export* goods responsible for five to six times as much pollution as 30 years ago. Seen that way, the rest of the world has been outsourcing pollution to the 24 high-income OECD nations.

So, have developed countries outsourced pollution by importing pollution-intensive products from the rest of the world? The answer depends in part on whether we think in absolute amounts of pollution or percentage changes. It also depends on whether we subtract the pollution embodied in exports from imports to calculate net pollution outsourced.

Table 2 presents both the absolute and percentage changes in trade and estimated pollution, calculated by comparing the starting and ending points of the lines in Figures 1 and 2. Pollution embodied in imports of the 24 high-income OECD countries from the rest of the world did grow, by 531 percent for PM_{10} and 830 percent for CO_2. Also, both grew faster than pollution embodied in exports from the high-income countries to the rest of the world, 429 percent and 477 percent. That looks like outsourcing pollution.

However, pollution embodied in trade also grew less steeply than overall trade. The dollar value of imports to the 24 high-income OECD countries grew 966 percent, while pollution embodied in those imports grew by less. Similarly, the dollar value of exports from the OECD grew by 678 percent, and pollution embodied in those exports grew less. The reason is that the mix of goods being imported *and* exported by the 24 high-income OECD countries shifted towards cleaner goods. That does not look like evidence of outsourcing pollution.

Figure 3
Pollution Intensity of Trade



Source: Author's calculations.
Note: Values calculated as the total predicted pollution in trade, from Figure 2, divided by the total volume of goods trade, from Figure 1 panel B.

Pollution Intensity Relative to Trade

Figure 3 depicts the pollution intensity of trade—estimated pollution content in Figure 2 divided by the dollar value of trade in Figure 1a. Or equivalently, it is a weighted average of industry-specific emissions intensities, where the weights are the dollar values of imports and exports. In each case, the values are based on the emissions intensities in the 2017 US National Emissions Inventory, so again, any change over time reflects only changes in the composition of goods being imported and exported, and nothing else. For both PM_{10} and CO_2 in Figure 3, the mix of manufactured goods imported by the 24 high-income countries is *less* polluting than the mix exported by those countries to the rest of the world. The solid lines lie below the dashed lines.

That finding may surprise readers who expect less developed countries to be producing the most polluting goods. What's the explanation? High-income countries have a comparative advantage in capital-intensive, high-skill industries that also happen to be relatively polluting. Thus, if polluting goods are traded, they are more likely to be exported by high-income countries, not imported. Moreover, a number of the most polluting industries—like petroleum refining, paper manufacturing, and cement—tend to be the most costly to transport long distances from local factor or product markets (Ederington, Levinson, and Minier 2005). That limits trade in those polluting goods, in either direction.

In addition to showing that the sample of high-income OECD nations imports a less polluting mix of goods than it exports, Figure 3 shows that the pollution intensity of neither imports nor exports is rising. For neither PM_{10} nor CO_2 have the OECD nations imported an increasingly polluting mix of goods. That does not sound like outsourcing.

For China's economy, which has developed rapidly and accounts for a large and growing fraction of trade with the OECD, pollution outsourcing also appears to be unimportant. For PM_{10} in Figure 3a, the mix of goods China sends to the 24 high-income countries was more polluting in 1989 than the mix that those countries imported or exported to the rest of the world. But by 2018, China was sending a less polluting mix of goods to the 24 OECD countries. For CO_2, China's mix of goods lies somewhere between that of the imports and exports of the 24 high-income countries.

Although the illustrative analysis in Figures 2 and 3 suggests little outsourcing of pollution, that conclusion comes with caveats. Trade flows are measured in US dollars, which has the effect of applying the same inflation adjustment to all industries. For example, energy price spikes will make it appear as though the pollution embodied in trade increases because the dollar value of energy-intensive shipments will increase—even if the physical volume does not. In addition, the analysis in Figures 2 and 3 classifies one particular set of 24 countries as developed, when in fact some of those 24 have grown more rapidly than others. Some of the countries outside the group of 24, like China, have themselves become more developed in the intervening years. Finally, the analysis uses the total emissions intensities reported by the US Environmental Protection Agency, which assume that all of the inputs to the products were also produced in the United States. The outsourcing depicted, therefore, describes the amount of pollution that would have been emitted in the 24 OECD countries if all imports and exports were produced domestically, *along with all inputs* to those imports and exports, using 2017 US technologies.

Congruence with Existing Research

Though this analysis is based on strong assumptions, the conclusions here corroborate earlier studies, which find that the composition of imports to rich countries has been shifting toward less polluting goods, not more polluting goods (Cole 2004; Brunel 2017; Levinson 2010). A nearly universal conclusion is that composition of production, including outsourcing and changing domestic consumption, plays a relatively small role, if any, in the ability of high-income economies to expand in recent decades without generating more pollution. Copeland, Shapiro, and Taylor (2022) summarize this research and add their own analysis, drawing the same conclusion: "Overall, this comparison across countries largely echoes the findings of previous country-specific studies—technology, rather than composition or scale effects, accounts for the largest share of the change in emissions."

That answers one important question: whether outsourcing accounts for environmental improvements in developed countries. Largely, it does not. But what about developing countries? The next section explores whether expanded trade has resulted in more pollution there.

Is Pollution Relocating to Developing Countries?

This question is complicated by the fact that countries have different production technologies. Some are more pollution-intensive than others, meaning they cause more pollution for the same output. If a country imports a product rather than producing it domestically, then total global pollution could increase or decrease, depending on which country employs the more pollution-intensive production technologies. Measuring the *net* pollution change due to trade requires knowing pollution intensities in both locations.

Asking whether just one single country outsources pollution in that way—increasing it by more abroad than it decreases at home—requires a slight modification to the input-output tables used in Leontief's (1970) calculation. Manufacturing an envelope using domestically produced paper involves more domestic pollution than manufacturing that same envelope using imported paper. Thus, to calculate pollution caused in the exporting country, we need to know the fraction of the paper used as an input that was itself also manufactured in that exporting country. In general, therefore, in addition to knowing pollution intensities in both countries, we need to know the share of each of the 360 NAICS industries that is produced domestically in both countries. That requires a lot of fine-grained data not universally available, but conceptually it is straightforward.

Asking whether a collection of relatively high-income countries like the OECD outsources pollution to another collection of countries makes that problem of imported intermediate inputs far trickier. In that case we do not care if the United States imports paper or envelopes from Canada, another country in the high-income group, but we do care if it imports them from a developing country. That means we must know not only the dollar value of each input required to manufacture a dollar of each output, but the dollar value of each input *from each other country*.

That information is called a multi-region input-output table. If the original country-specific input-output table covers 360 industries, then examining outsourcing by 24 OECD countries to 100 rest-of-the-world countries would require an input-output table with $(360 \times 124) \times (360 \times 124)$ entries. That is nearly two billion data points. The next step would require each of the 124 countries to report emissions intensities for each of those 360 industries. So far, only a few countries report emissions in anywhere close to that detail. Those that do tend to use different definitions of industries and pollution intensities.

Several organizations have assembled versions of multi-region input-output tables.[8] Most aggregate manufacturing industries into only a few dozen sectors. None has industry detail comparable to the US National Emissions Inventory used to draw Figures 2 and 3. As a consequence, those multi-region input-output tables cannot distinguish between imports of paper, which outsource pollution, and

[8] See, for example, the World Input-Output Database (Timmer et al. 2015), Exiobase (Stadler et al. 2018), and Eora26 (Lenzen et al. 2013). The website http://environmentalfootprints.org describes those and others (accessed 4/14/2022).

imports of envelopes, which may or may not, depending on where the paper for those envelopes is produced.

Copeland, Shapiro, and Taylor (2022) use one of those multi-region input-output tables, the World Input Output Dataset, to answer this outsourcing question. They use an accounting exercise similar to that behind Figure 2 here, but different in one key way. In their version (see their Figure 5), the solid line corresponding to pollution caused by imports to high-income countries is measured using pollution intensities for the countries doing the exporting. It approximates how much pollution was caused in developing countries as a consequence of their exports. Recall that in Figure 2, the solid line corresponding to pollution embodied in OECD imports is measured using the US National Emissions inventory. That approximates the pollution that would have occurred in the OECD had those goods been produced domestically rather than imported. It is an important distinction. If the same goods were being traded back and forth, in the same amounts, Copeland, Shapiro, and Taylor (2022) would show net outsourcing of pollution, while Figure 2 would not.

In the Copeland, Shapiro, and Taylor (2022) figure, pollution embodied in rich-country imports is larger than pollution embodied in rich-country exports (equivalent to the sold line being above the dashed one in Figure 2). That is consistent with richer countries having stricter pollution standards and using cleaner production technologies, even for the same goods. The gap between the two lines also grows over time, although it is not clear whether that represents changes to the mix of goods imported and exported, changes to the emissions intensities in rich and poor countries, or as the authors note, simply "changes in the scale of net trade flows."

Copeland, Shapiro, and Taylor (2022) conclude, based on their version of Figure 2, that "rich countries are increasingly outsourcing pollution." Although that sounds contradictory to most of the prior results and to the illustration in the previous section, they are asking a different question. They ask how much larger pollution is in developing countries as a consequence of exports. Figure 2 asks how much more pollution there would have been in developed countries without those imports. It should not be surprising that the answers are different.

Another important distinction between the analyses involves industry aggregation. Copeland, Shapiro, and Taylor (2022) and others (like Peters et al. 2011) that reach similar conclusions use multi-region input-output tables with only a few dozen categories of manufacturing industries. The 2013 World Input Output Dataset had only 24 different categories of imports, 14 of which are manufactured goods. In these kinds of datasets, paper mills and envelope manufacturers are combined with printers and publishers into one industry, with one pollution intensity. It is possible that this aggregation, though necessary to compare trade among multiple countries with incomplete or incompatible data, either exaggerates or understates the pollution embodied in trade.

Finally, because these studies estimate poor-country emissions intensities to be larger than rich-country intensities, when the scale of trade grows they show net pollution imports to be increasing. If nothing changes except that imports and exports both double, their measure of the pollution embodied in net imports would also double. Whether to call that "outsourcing pollution" seems debatable, and more about semantics than economics.

That leaves our third question. Have the environmental regulations enacted by developed countries caused them to increase imports of polluting goods?

Do Environmental Regulations in High-Income Countries Cause Outsourcing of Pollution?

In hindsight, a sensible research agenda might have started with the answer to our title question: Is there outsourcing of pollution? And then, if the answer was "yes," the next step would be to examine the causes of this outsourcing, one of which might include more stringent environmental regulations in developed countries. But chronologically, that is not how this research has progressed. Economists' study of pollution outsourcing began with the harder question: whether strict environmental standards lead to outsourcing of pollution.

One of the very first studies of whether regulations cause outsourcing illustrates the difficulties in determining that cause-and-effect relationship. Weighing into the debate about the environmental consequences of the North American Free Trade Agreement (NAFTA) enacted in 1994, Grossman and Krueger (1993) asked whether industries facing higher pollution abatement costs in the United States were more likely to be outsourced to Mexico. They regressed US imports from Mexico in each of 136 industries on those industries US pollution abatement costs and other characteristics. The pollution cost coefficients are small and statistically insignificant, indicating that industries facing higher pollution costs in the United States were not more likely to be imported from Mexico. The researchers concluded that US "environmental regulations and enforcement . . . play at most a minor role" in outsourcing.

However, in some specifications, Grossman and Krueger (1993) found that goods with higher pollution costs when manufactured in the United States were statistically significantly *less* likely to be imported from Mexico. Nobody concludes from that result that the United States is a pollution haven for certain Mexican industries. More likely, this cross-section empirical work, with a mere 136 observations, has trouble accounting for omitted variables and the endogeneity of the regulations. Places that find themselves with worse pollution, all else equal, will want to set stricter standards. Indeed, the Clean Air Act in the United States requires places with the worst air quality to have the strictest pollution regulations. If the poor air quality arises as a result of some local comparative advantage for polluting industries, like abundant natural resources or access to transport, the places with strictest standards will also be the places where polluting industries locate and consequently where the pollution is worst. In the US-Mexico context, it will appear as though Mexico is outsourcing pollution to the United States.

Properly answering whether pollution regulations cause outsourcing of pollution requires two kinds of data and a methodology that can be interpreted in cause-and-effect terms. The first data requirement is a measure of the stringency of pollution regulations, which is not easy to come by. Grossman and Krueger (1993) and many subsequent researchers used a survey of US manufacturers called the Pollution Abatement Costs and Expenditures Survey. But that survey only covered the United States, and it was last conducted in 2005. Other researchers have used surveys of business managers, such as the WEF Executive Opinion Survey (Kellenberg 2009; Wagner and Timmins 2009). But that survey solicits perceptions, and executives from polluting industries are more likely to perceive strict regulations (Kalamova and Johnstone 2012). Still others rely on counts of the number of regulations or on econometric estimates of the inefficiencies caused by regulations. In the future, if more places enact regulations to put a price on carbon, perhaps researchers may be able to use that price as a measure of regulatory stringency going forward. But in the meantime, measuring the strictness of environmental regulations across varying pollutants remains logistically and conceptually difficult (Brunel and Levinson 2016).

The second data requirement is a measure of how regulatory stringency across industries changes over time. Grossman and Krueger (1993) showed that industries facing high pollution regulation costs in the United States are not more likely to be imported from Mexico. However, it is possible that industries facing larger *increases* in costs may be more likely to increase their imports, consistent with an outsourcing story. Why the difference? Comparing changes in stringency to changes in imports holds location-specific differences constant, effectively controlling for omitted characteristics of countries and industries that are fixed over time, like natural resources or good transport. This second requirement magnifies the difficulty associated with the first. We need data on changes over time in a difficult-to-measure concept.

The methodological challenge to answering whether regulations cause outsourcing involves the fact that changes in regulations are endogenous. Locations that find themselves experiencing increasing pollution, due to a locally expanding polluting industry, are more likely to tighten their standards. That tightening could create a positive correlation between the presence of polluting industries and strict antipollution regulations, spuriously making it seem as though strict regulations attract polluting businesses. A typical solution would be to seek an instrumental variable, something correlated with changes in pollution regulations but uncorrelated with outsourcing except through the effect of those regulations. That compounds the difficulty of properly answering the question. We need an instrumental variable to proxy for changes over time in a difficult-to-measure concept. That sounds daunting, and it is.

One new paper that comes close to overcoming these data and methodological hurdles is Cherniwchan and Najjar (2022). They study Canada's rule that required every major city and town to achieve at least a minimum level air quality. Affected manufacturers located in places failing to meet the standard became subject to new, more strict pollution rules. The measure of stringency is easily quantifiable, if blunt:

some places meet the standard, others do not. It is a measure that changes over time as cities move into or out of compliance. It is plausibly exogenous, given that the standard was set nationally. Cherniwchan and Najjar (2022) find that targeted factories decreased exports by more than 20 percent. That does not, however, mean that polluting production was outsourced to developing countries. Production could have moved to other Canadian provinces or to the United States, or been matched by a decline in use of those products.

In general, recent surveys of research into this third question have a consensus. Dechezleprêtre and Sato (2017), Cole, Elliott, and Zhang (2017), Jakob (2021), and Caron (2022) all find that pollution regulations impose costs that are relatively small shares of total costs for most industries and just one of many such costs that affect import decisions. In 2005, when the United States last surveyed manufactures about this, pollution abatement costs amounted to less than one-half of a percent of the value of manufacturing shipments, ranging from less than 0.1 percent for textiles to a max of 1.1 percent for primary metals (Bureau of the Census 2008). Research that appropriately controls for omitted variables and endogenous regulations can sometimes find statistically significant effects of regulations on trade, but when it does, the estimated magnitudes are small.

The review articles also note that for some industries that are particularly pollution-intensive, and easily imported or exported, regulations may meaningfully cause outsourcing of pollution. As an example, Tanaka, Teshima, and Verhoogen (2022) study one industry, battery recycling, and one regulation, the US standard for airborne lead. They show convincingly that when the regulation tightened in 2009, air quality improved in neighborhoods near US battery recyclers and degraded near Mexican battery recyclers, and that US exports of used batteries to Mexico increased. That seems like a clear case study providing rare empirical evidence of regulation-induced outsourcing of pollution, but it does not appear to represent a general pattern.

In sum, answering this third question—whether environmental regulations cause pollution outsourcing—is profoundly difficult. In the absence of regulations that price pollution explicitly, measuring regulatory stringency poses a conceptual challenge. Few researchers who try to answer the question have panel data allowing them to examine the effect of changes in stringency over time on changes in trade flows. Even fewer have managed to deploy a convincing instrumental variable to control for the fact that places with growing pollution problems enact the strictest rules. Summarizing the research in this area, the World Bank's annual flagship publication, the *World Development Report*, concludes that "strict environmental regulation of polluting industries has not led to large relocations to countries with less strict standards" (World Bank 2020, p. 125).

Conclusion

For those who are concerned about the potential risks of outsourcing pollution from high-income countries to the rest of the world, an obvious workaround is border taxes, or tariffs, based on the pollution embodied in imports. Such a policy would have two objectives.

First, border pollution taxes would seek to prevent outsourcing pollution. However, if the main thrust of the research described here is correct, even without border taxes there has been limited meaningful outsourcing. In that case, border taxes would be protecting domestic industries from an imagined threat. In political terms, perhaps those border taxes are a necessary chip in bargaining over domestic environmental policy rules (Jakob et al. 2022). In fact, there is some evidence that tariffs are already higher on goods facing more costly domestic environmental standards (Ederington and Minier 2003), although Shapiro (2021) shows that tariffs are on average lower for more polluting imports. In this issue, border taxes are discussed in greater depth in the paper by Clausing and Wolfram.

Second, border pollution taxes would be a way for higher-income countries to encourage exporting countries in the rest of the world to set more stringent environmental standards of their own. For local pollutants like PM_{10}, the justification for high-income countries to exert this kind of pressure on exporters in the rest of the world may arise from concerns about environmental quality in those countries. For global pollutants like CO_2, the justification would be preventing leakage.

If the question is whether developed countries cleaned their own environments by importing polluting goods, then the answer seems to be "no." Imports and exports to and from developed countries have grown rapidly. But the mix of goods imported has not tilted disproportionately towards relatively more polluting industries. In fact, over the past three decades that mix has shifted towards cleaner industries, not dirtier.

If the question is whether developed country imports have resulted in more pollution elsewhere, the answer seems to be "it depends." Detailed industry-specific emissions intensities are not available for most countries. As a result, existing multiregion input-output tables use broad classifications that blur important distinctions between polluting and clean industries in the same sector. Attempts to use those do find that emissions intensities for those broad classifications are higher in poor countries than in rich ones, which means that even balanced trade between poor and rich countries will show more pollution caused by developed country imports than exports, even for the same goods. And proportional growth in imports and exports will show that gap to be increasing. Whether we should label that "outsourcing pollution" is unclear.

Finally, if the question is whether regulations in developed countries cause polluting industries to relocate to pollution havens, the answer is that identifying that cause-and-effect relationship is tricky. For specific, geographically mobile, and pollution-intensive industries, examples can be demonstrated. But in general, the balance of the evidence to date does not find statistically or economically significant evidence of regulations causing outsourcing. For all the talk of outsourcing pollution in the media and politics, there is surprisingly little empirical evidence that high-income regions increasingly and disproportionally import products of the most polluting sectors.

■ *Many thanks to Claire Brunel, Jevan Cherniwchan, Josh Ederington, Xuerui Mei, Yağmur Menzilcioğlu, and Joe Shapiro for answering my questions and offering helpful suggestions.*

References

Böhringer, Christoph, Carolyn Fischer, Knut Einar Rosendahl, and Thomas Fox Rutherford. 2022. "Potential Impacts and Challenges of Border Carbon Adjustments." *Nature Climate Change* 12: 22–29.

Brunel, Claire. 2017. "Pollution Offshoring and Emission Reductions in EU and US Manufacturing." *Environmental and Resource Economics* 68 (3): 621–41.

Brunel, Claire, and Arik Levinson. 2016. "Measuring the Stringency of Environmental Regulations." *Review of Environmental Economics and Policy* 10 (1): 47–67.

Bureau of the Census. 2008. "Pollution Abatement Costs and Expenditures: 2005 (MA200-2005)." US Department of Commerce. http://www.epa.gov/environmental-economics/pollution-abatement-costs-and-expenditures-2005-survey (accessed May 4, 2023).

Bureau of Economic Analysis. 1947-2017. "GDPbyIndGO_1947-2017.xlsx" United States Department of Commerce. https://www.bea.gov/sites/default/files/2018-04/GDPbyInd_GO_1947-2017.xlsx (accessed May 7, 2022).

Campbell, Erin, Anne McDarris, and Billy Pizer. 2021. *Border Carbon Adjustments 101*. Washington, DC: Resources for the Future.

Caron, Justin. 2022. "Empirical Evidence and Projections of Carbon Leakage: Some, but Not Too Much, Probably." In *Handbook on Trade Policy and Climate Change*, edited by Michael Jakob, 58–74. Cheltenham, UK: Edward Elgar Publishing.

Cherniwchan, Jevan, and Nouri Najjar. 2022. "Do Environmental Regulations Affect the Decision to Export?" *American Economic Journal: Economic Policy* 14 (2): 125–60.

Cole, Matthew A. 2004. "US Environmental Load Displacement: Examining Consumption, Regulations and the Role of NAFTA." *Ecological Economics* 48 (4): 439–50.

Cole, Matthew A., Robert J. R. Elliott, and Toshihiro Okubo. 2014. "International Environmental Outsourcing." *Review of World Economics* 150 (4): 639–64.

Cole, Matthew A., Robert J. R. Elliott, and Liyun Zhang. 2017. "Foreign Direct Investment and the Environment." *Annual Review of Environment and Resources* 42: 465–87.

Copeland, Brian R., Joseph S. Shapiro, and M. Scott Taylor. 2022. "Globalization and the Environment." In *Handbook of International Economics*, Vol. 5, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 61–146. Amsterdam: North-Holland.

Cronkite, Walter. 1980. "To Save Our Industry and the Environment" *New York Times* October 14, p.A18. https://www.sciencedirect.com/handbook/handbook-of-international-economics/vol/5/suppl/C.

Dechezleprêtre, Antoine, and Misato Sato. 2017. "The Impacts of Environmental Regulations on Competitiveness." *Review of Environmental Economics and Policy* 11 (2): 183–206.

Ederington, Josh, Arik Levinson, and Jenny Minier. 2005. "Footloose and Pollution-Free." *Review of Economics and Statistics* 87 (1): 92–99.

Ederington, Josh, and Jenny Minier. 2003. "Is Environmental Policy a Secondary Trade Barrier? An Empirical Analysis." *Canadian Journal of Economics* 36 (1): 137–54.

European Commission, Joint Research Centre (EC-JRC)/Netherlands Environmental Assessment Agency (PBL). 1970–2018. "Emissions Database for Global Atmospheric Research (EDGAR)." https://edgar.jrc.ec.europa.eu/overview.php?v=50_AP (accessed December 21, 2021).

Grossman, Gene M., and Alan B. Krueger. 1993. "Environment Impacts of a North American Free Trade Agreement." In *The Mexican-U.S. Free Trade Agreement*, edited by Peter M. Garber, 13–56. Cambridge, MA: MIT Press.

Hausman, Daniel M., and Michael S. McPherson. 2006. *Economic Analysis, Moral Philosophy and Public Policy*. 2nd ed. New York: Cambridge University Press.

Jakob, Michael. 2021. "Climate Policy and International Trade—A Critical Appraisal of the Literature." *Energy Policy* 156: 112399.

Jakob, Michael, Stavros Afionis, Max Åhman, Angelo Antoci, Marlene Arens, Fernando Ascensão, Harro van Asselt et al. 2022. "How Trade Policy Can Support the Climate Agenda." *Science* 376 (6600): 1401–03.

Kalamova, Margarita, and Nick Johnstone. 2012. "Environmental Policy Stringency and Foreign Direct Investment." In *A Handbook of Globalisation and Environmental Policy*, 2nd ed, edited by Frank Wijen, Kees Zoeteman, Jan Pieters, and Paul van Seters, 34–56. Cheltenham, UK: Edward Elgar Publishing.

Kellenberg, Derek. 2009. "An empirical investigation of the pollution haven effect with strategic environment and trade policy" Journal of International Economics 78 (2): 242–55.

Lenzen, Manfred, Daniel Moran, Keiichiro Kanemoto, and Arne Geschke. 2013. "Building Eora: A Global Multi-regional Input–Output Database at High Country and Sector Resolution." *Economic Systems Research* 25 (1): 20–49.

Leontief, Wassily. 1970. "Environmental Repercussions and the Economic Structure: An Input–Output Approach." *Review of Economics and Statistics* 52 (3): 262–71.

Leonard, H. Jeffrey. 1988. *Pollution and the Struggle for the World Product: Multinational Corporations, Environment, and International Comparative Advantage.* Cambridge, UK: Cambridge University Press.

Levinson, Arik. 2003. "Environmental Regulatory Competition: A Status Report and Some New Evidence." *National Tax Journal* 56 (1.1): 91–106.

Levinson, Arik. 2009. "Technology, International Trade, and Pollution from US Manufacturing." *American Economic Review* 99 (5): 2177–92.

Levinson, Arik. 2010. "Offshoring Pollution: Is the United States Increasingly Importing Polluting Goods?" *Review of Environmental Economics and Policy* 4 (1): 63–83.

Levinson, Arik. 2023. "Replication data for: Are Developed Countries Outsourcing Pollution?" American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E191621V1.

Oates, Wallace E., and Robert M. Schwab. 1996. "The Theory of Regulatory Federalism: The Case of Environmental Management." In *The Economics of Environmental Regulation*, edited by Wallace E. Oates, 319–31. Brookfield, VT: Edward Elgar Publishing.

Organization for Economic Cooperation and Development (OECD). "GDP Implicit Price Deflator in United States [USAGDPDEFQISMEI])." Retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/USAGDPDEFQISMEI (accessed December 14, 2021).

Organization for Economic Cooperation and Development (OECD). 1960–2022. "Gross domestic product (GDP)." https://doi.org/10.1787/dc2f7aec-en (accessed March 31, 2022).

Pashigian, B. Peter. 1985. "Environmental Regulation: Whose Self-Interests Are Being Protected?" *Economic Inquiry* 23 (4): 551–84.

Peters, Glen P., Jan C. Minx, Christopher L. Weber, and Ottmar Edenhofer. 2011. "Growth in Emission Transfers via International Trade from 1990 to 2008." *Proceedings of the National Academy of Sciences* 108 (21): 8903–08.

Pierce, Justin R., and Peter K. Schott. 2012. "Concording U.S. Harmonized System Codes over Time." *Journal of Official Statistics* 28 (1): 53–68.

Shapiro, Joseph S. 2021. "The Environmental Bias of Trade Policy." *Quarterly Journal of Economics* 136 (2): 831–86.

Sierra Club. 2022. "Carbon Dumping Fees: A Tool to Close the Climate Loophole." www.sierraclub.org/sites/www.sierraclub.org/files/carbon-dumping-fee-climate-trade.pdf (accessed May 31, 2022).

Stadler, Konstantin, Richard Wood, Tatyana Bulavskaya, Carl-Johan Södersten, Moana Simas, Sarah Schmidt, Arkaitz Usubiaga et al. 2018. "EXIOBASE 3: Developing a Time Series of Detailed Environmentally Extended Multi-regional Input-Output Tables." *Journal of Industrial Ecology* 22 (3): 502–15.

Tanaka, Shinsuke, Kensuke Teshima, and Eric Verhoogen. 2022. "North-South Displacement Effects of Environmental Regulation: The Case of Battery Recycling." *American Economic Review: Insights* 4 (3): 271–88.

Timmer, Marcel P., Erik Dietzenbacher, Bart Los, Robert Stehrer, and Gaaitzen J. de Vries. 2015. "An Illustrated User Guide to the World Input–Output Database: The Case of Global Automotive Production." *Review of International Economics* 23 (3): 575–605.

United Nations Statistics Division, UN COMTRADE. 1989–2018. "International Merchandise Trade Statistics." http://comtrade.un.org/ (accessed December 18, 2021).

Wagner, Ulrich, and Christopher Timmins. 2009. "Agglomeration effects in foreign direct investment and the pollution haven hypothesis" *Environmental and Resource Economics* 43(2): 231–56.

World Bank. 2020. *World Development Report 2020: Trading for Development in the Age of Global Value Chains.* Washington, DC: World Bank.

Yang, Yi, Wesley W. Ingwersen, Troy R. Hawkins, Michael Srocka, and David E. Meyer. 2017. "USEEIO: A New and Transparent United States Environmentally-Extended Input-Output Model." *Journal of Cleaner Production* 158: 308–18.

Journal of Economic Perspectives—Volume 37, Number 3—Summer 2023—Pages 111–136

Think Globally, Act Globally: Opportunities to Mitigate Greenhouse Gas Emissions in Low- and Middle-Income Countries

Rachel Glennerster and Seema Jayachandran

Most environmental externalities are, to some degree, local. Conventional air pollutants in India or China primarily affect people living in those countries, even if a small part of the pollution reaches the United States.

Climate change is unusual in that it is a truly global externality: A ton of CO_2 emissions, or emissions of other greenhouse gases like methane and nitrous oxide, adds to the Earth's atmospheric levels of these gases, regardless of whether it was emitted in the United States or Spain or Kenya. The global nature of the problem makes it harder to solve, because a large share of the costs of worsening climate change are borne by those outside the borders of the emitting country. Moreover, while climate change will touch all countries, the effects will be uneven. Low- and middle-income countries are expected to be hardest hit, partly because of their geography—many are in already-hot regions, so warming leads to dangerously high temperatures—but also because livelihoods are more fragile, and the options for coping with climate change, from air conditioners to flood barriers, are often out of their financial reach.

However, the global nature of this externality also offers an opportunity. Because global atmospheric CO_2 decreases by the same amount whether a ton of emissions is reduced in the United States or Spain or Kenya, countries (or other actors) who want to reduce emissions of carbon dioxide and other greenhouse gases can seek

■ *Rachel Glennerster is Associate Professor of Economics, University of Chicago, Chicago, Illinois. Seema Jayachandran is Professor of Economics and Public Affairs, Princeton University, Princeton, New Jersey. Their email addresses are rglennerster@uchicago.edu and jayachandran@princeton.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.3.111.

the most cost-effective ways to reduce emissions anywhere in the world, not just within their borders.

This approach has not yet taken hold: The vast majority of mitigation takes place within the funders' borders, resulting in mitigation activity that is highly concentrated in a few countries, which is unlikely to be efficient. Climate financing for mitigation in Western Europe is $105 billion compared to $30 million in South Asia (Buchner et al. 2021), despite substantially higher population and land area in South Asia.[1] This imbalance exists despite provisions in the Kyoto Protocol (the 1997 international climate treaty) and the Paris Agreement (the 2015 climate treaty) for countries to meet their emissions targets by funding mitigation activity abroad.

In this article, we will argue that many of the most cost-effective opportunities for mitigation—that is, reduction in atmospheric CO_2 levels—are likely to be in low- and middle-income countries. The implication of this reasoning is that high-income countries, as well as multilateral agencies and philanthropists, could and should be tapping opportunities in low- and middle-income countries to achieve more mitigation for a given level of spending.

The Paris Agreement set a target of limiting global temperature rise to less than 2.0°C above pre-industrial levels, while positioning 1.5°C of warming as the preferable goal. Climate scientists have warned of dire consequences if the more ambitious 1.5°C target is not met, because any additional warming brings the planet closer to catastrophic tipping points such as the collapse of Greenland's ice cap (Hoegh-Guldberg et al. 2019; Armstrong McKay et al. 2022). Under the Paris Agreement, countries set so-called "nationally determined contributions," which are their targets for greenhouse gas emissions, but there are large gaps between these current targets and the emissions cuts required to limit global warming to 1.5°C by 2030 (UN Climate Change 2021).

The UN Environment Programme (2021) estimated that staying below 1.5°C will require a 55 percent reduction in greenhouse gas emissions—or almost twice as much as the 30 percent cut needed for 2°C of warming (UNEP 2021). What is needed to double, or at least dramatically increase, the amount of mitigation? Larger financial commitments from the world's governments are surely needed, but we must also search for and fund the most cost-effective sources of mitigation. If we could shift funding to mitigation projects that are twice as cost-effective as what

[1] Mitigation funding is more concentrated than greenhouse gas emissions. Emissions of greenhouse gases other than CO_2 are converted into their CO_2-equivalent (CO2e) weight. Western Europe emits 3.1 billion metric tons of CO2e per year, which is 35 percent less than the 4.8 billion metric tons a year emitted in South Asia (Ritchie, Roser, and Rosado 2020a), but as we discuss below, emissions and mitigation being geographically aligned should not be the objective. Western Europe refers to Andorra, Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Liechtenstein, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, San Marino, Slovenia, Spain, Sweden, Switzerland, United Kingdom, and Vatican City, and South Asia refers to Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan, and Sri Lanka to match the categorization used by Buchner et al. (2021).

we are currently doing, that would effectively double the impact of the funding for carbon mitigation.

Who should pay for carbon mitigation and where that mitigation is most efficiently conducted are two separate questions. Unfortunately, focus on the "who should pay" question has often obscured clear-eyed discussion of "where is the reduction most efficiently conducted," which is the primary subject of our article.

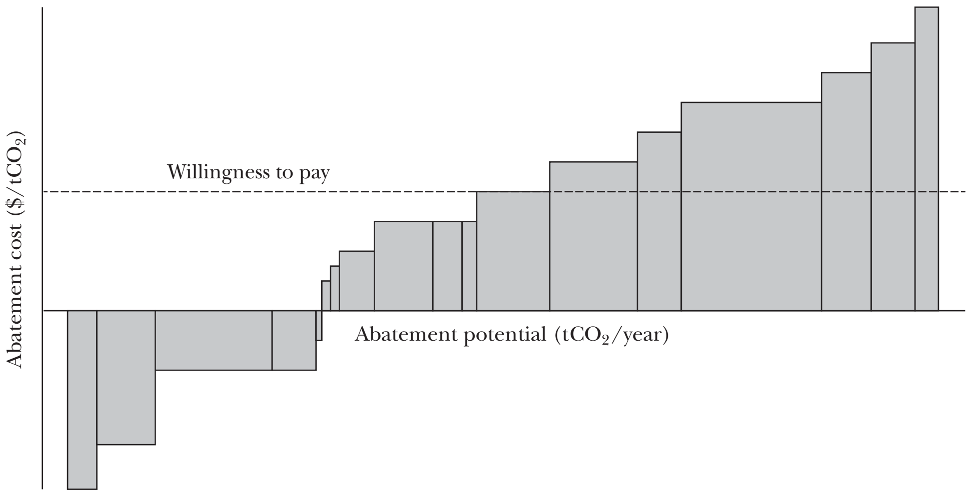
In this essay, we begin by introducing an "abatement cost curve," which shows that there are likely to be a range of ways for reducing carbon emissions across sectors and countries. We then turn to four reasons that climate mitigation in low- and middle-income countries is economically attractive: (1) many of the easiest and cheapest options to reduce greenhouse gas emissions have already been tapped in high-income countries; (2) immobile inputs that are used in mitigation projects, namely land and labor, are cheaper in low- and middle-income countries; (3) it is cheaper to build mitigation into new infrastructure in low- and middle-income countries than to retrofit existing infrastructure in high-income countries; and (4) general equilibrium considerations imply that a geographically balanced approach to mitigation across both high-income countries and low- and middle-income countries is needed. We also discuss forces pushing in the other direction—that is, reasons that carbon mitigation might be cheaper in high-income countries.

Although our focus in this essay is about where carbon mitigation should be conducted, the question of who shall pay cannot be sidestepped. High-income countries are responsible for most of the carbon already emitted as a result of human activity. Even today, after growth of economies and carbon emissions in places like China, India, and Indonesia, the richest 10 percent of the world is still responsible for about half of the world's greenhouse gas emissions. In addition, high-income countries have greater resources to pay for mitigation. However, a rapidly rising share of global emissions going forward will come from middle-income countries, and thus, policymakers in high-income countries argue that these countries should also bear some cost of and responsibility for action on their part. Here, we will not seek to advance the well-rehearsed arguments about how much high-income countries should pay for mitigation. However, we will explore some of the broader ethical and practical policy arguments for how actions and/or payments by high-income countries to mitigate carbon emissions might, on grounds of addressing a major environmental hazard more effectively, be usefully focused on low- and middle-income countries.

The Abatement Cost Curve

There are myriad greenhouse-gas-emitting human activities, which means that there are also myriad opportunities to reduce emissions. An abatement cost curve depicts the various available opportunities to reduce, or abate, greenhouse gas emissions, depicting the cost and size of each opportunity. We will refer to the abatement

Figure 1
An Abatement Cost Curve for Greenhouse Gas Emissions



Source: Authors' illustration.
Note: An illustrative cost curve depicting the possible ways to reduce (or abate) greenhouse gas emissions. Each bar represents a different way to reduce emissions. A bar's width is how much emissions could be reduced in that way, and its height is the cost per ton to do so. The bars are ordered from lowest- to highest-cost opportunities.

cost curve to help explain the different reasons that there are low-cost mitigation opportunities in low- and middle-income countries.

In the abatement cost curve in Figure 1, each bar represents an activity that could reduce greenhouse gas emissions, such as reforesting deforested land or adopting hybrid vehicles (to displace gasoline-engine vehicles). While typically these activities are not delineated by location on an abatement cost curve, for our purposes it is useful to sometimes think of separate bars for the same activity but in different locations, such as reforestation in high-income versus in low- and middle-income countries.

The units along the horizontal axis of the figure are the feasible amount of abatement, often expressed in metric tons of CO_2-equivalent (tCO2e) per year—so that actions to reduce other greenhouse gases like methane and nitrous oxide can be shown on a common scale with carbon emissions. The width of a bar represents how large an opportunity that activity represents: wider bars could achieve more of the mitigation that the world needs. The vertical axis is the cost per metric ton of carbon-equivalent for the activity. The taller the bar, the more expensive that abatement cost. The bars are ordered by their height, with the lowest cost options on the left. The graph stops at some point on the right, ignoring other, more expensive options beyond those depicted. Some bars in our stylized abatement cost curve have a negative cost, which means they would save people money if they undertook them. Replacing incandescent light bulbs with LED bulbs is a classic example: although it requires some upfront money to buy and install the

Table 1
Breakdown of Greenhouse Gas Emissions by Sector, Separated by Income Group

Sector	*High-income countries*	*Upper-middle-income countries*	*Lower-middle-income countries*	*Low-income countries*
Electricity and heat	33.0%	38.4%	23.1%	2.6%
Transport	24.5%	10.4%	10.2%	5.1%
Manufacturing and construction	9.6%	17.1%	10.8%	4.2%
Buildings	9.1%	4.9%	5.4%	3.4%
Agriculture	7.4%	9.4%	18.9%	47.9%
Aviation and shipping	5.9%	1.3%	0.7%	0.5%
Industry	5.3%	7.8%	5.2%	2.7%
Fugitive emissions	5.1%	8.3%	5.4%	3.3%
Waste	2.6%	3.2%	4.4%	4.6%
Other fuel combustion	1.1%	1.4%	1.2%	0.8%
Land-use change and forestry	–3.5%	–2.1%	14.8%	24.8%
Total	100%	100%	100%	100%

Source: Ritchie, Roser, and Rosado (2020a).
Note: Fugitive emissions are leaks from pipelines, wells, appliances, storage tanks, pipelines, wells, or other equipment.

LEDs, they are more energy-efficient, so the savings in electricity bills more than covers the initial investment cost.

An abatement cost curve is a supply curve, and to see the equilibrium on the graph, one needs to overlay the demand curve. If the world were willing to pursue all projects up to a certain cost, the demand curve would be a horizontal line at that level—the willingness to pay—on the vertical axis. All the bars shorter than the willingness to pay (those to the left of where willingness to pay intersects a bar) would be implemented in equilibrium.

The abatement curve depicted here is a global one, but if we created different curves for different groups of countries, the bars for a particular activity would differ in width across countries because the composition of emissions varies across countries. Table 1 compares greenhouse gas emissions by sector for categories of countries with levels of per capita income, as defined by the World Bank for 2023: high-income countries (>$13,205 in per capita GDP), upper-middle-income countries ($4,256 to $13,205), lower-middle-income countries ($1,086 to $4,255), and lower-income countries (<$1,085). The differences are stark. One-third of emissions in high-income countries comes from electricity and heating, compared to less than 3 percent in low-income countries. Transport, including aviation and shipping, is a much bigger contributor in high-income than in middle-income countries, and especially than in low-income countries where car ownership and usage are much lower. Meanwhile, land use change and forest loss are a major component of emissions in low- and lower-middle-income countries, but not elsewhere. Similarly, agriculture is a major carbon-emitting sector only in low- and lower-middle-income countries.

We now turn to the four reasons for low-cost options in low- and middle-income countries.

Reason #1: The Easiest and Cheapest Options Have Already Been Tapped in High-Income Countries

One reason for the mitigation bargains in low- and middle-income countries is that some cheap opportunities that are already being pursued in high-income countries remain untapped in low- and middle-income countries. Here, we discuss three economic drivers to explain this pattern.

Differences in Willingness to Pay

We would expect households and governments in low- and middle-income countries to have a lower willingness to pay for carbon mitigation than their counterparts in high-income countries. Governments and citizens in low- and middle-income countries will, understandably, prioritize spending on basic needs and promoting economic growth, which could raise their relatively low standard of living, over contributing to the global public good of protecting the planet's health.

Survey data suggest that environmental protection is a tougher tradeoff to make in low- and middle-income countries. The World Values Survey asks a representative sample of adults in many countries about their social, economic, political, and cultural values every few years. One question on the survey asks respondents which of the two statements better reflects their view:

1. Protecting the environment should be given priority, even if it causes slower economic growth and some loss of jobs.

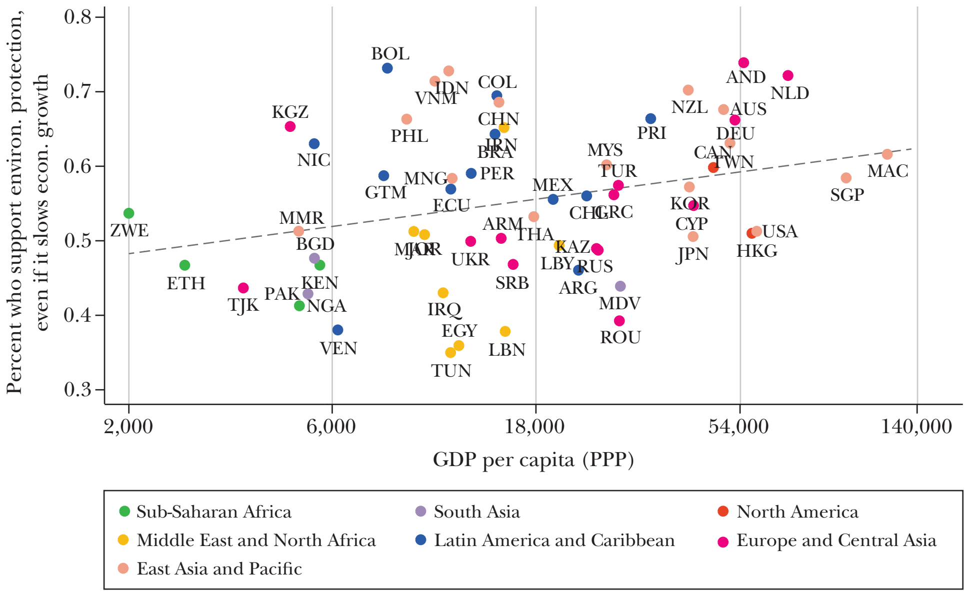
2. Economic growth and creating jobs should be the top priority, even if the environment suffers to some extent.

Figure 2 plots the average responses by country. The vertical axis is the proportion of people choosing the first, pro-environment position. The horizontal axis is the country's real GDP per capita. While there is a lot of variation unexplained by GDP per capita, support for protecting the environment, even when it slows growth and causes job loss, indeed tends to be higher as income rises.

In the context of the abatement curve in Figure 1, if willingness to pay has historically been lower outside of the high-income countries, then the marginal opportunity for abatement in low- and middle-income countries will be lower-cost than its counterpart in high-income countries. Or to put it another way, some of the abatement opportunities that remain in low- and middle-income countries are less expensive than the best remaining opportunity in high-income countries.

One example is decommissioning coal-fired power plants in favor of cleaner power generation. The share of electricity generation from coal has fallen steadily in high-income countries over the past few decades but has grown, or at least not fallen, in middle-income countries. In 2021, 19 percent of electricity generation was from coal in high-income countries versus over 45 percent in middle-income

Figure 2
Cross-Country Comparison of Attitudes about Environment-Growth Tradeoffs



Source: Data are from the World Values Survey, Wave 7, collected between 2017 and 2022.
Note: GDP per capita is for the specific country's survey year and is based on purchasing parity power and expressed in 2022 US dollars.

countries (Ritchie, Roser, and Rosado 2020b). In the US economy, this transition occurred partly because natural gas, a cleaner fossil fuel, became cheaper, but also because renewable alternatives became cheaper. Funding at least the start of a transition away from coal in low- and middle-income countries is a low-cost abatement opportunity, in part because of the dirty starting point. Relatedly, funding wind or solar power might be a more cost-effective way to reduce emissions in India—where 74 percent of electricity generation uses coal—than in the United States, partly because renewables would typically be displacing a dirtier energy source in India, leapfrogging over natural gas.

However, differences in willingness to pay across countries still leave a puzzle: Why would any country not take advantage of any existing negative-cost opportunities for abatement? One possibility is that the cost estimates are wrong; for example, such estimates often omit policy implementation costs, like those related to new regulatory apparatuses that are needed, as well as political costs. Nonetheless, there very likely are profitable opportunities that have not been pursued. Two answers to the puzzle of why they have not been pursued are limited access to capital and limited regulatory capacity, as we elaborate in the rest of this section. Both of these explanations also suggest that low- and middle-income countries will have more negative-cost abatement options than high-income countries.

Limited Access to Capital

Mitigation opportunities often entail upfront costs, with the savings that cover those costs accruing over time. The cost of borrowing is higher for low- and middle-income country governments, mostly because the perceived risk of default is higher. Similarly, households in low- and middle-income countries have less access to capital because the financial sector is less developed.

Methane capture from landfills is an example where upfront spending could be recouped over time—in this case, by means of an income stream. Landfills generate methane, a greenhouse gas, as organic material decomposes. If no measures are taken to manage the methane, it escapes the landfill and adds to atmospheric greenhouse gas levels. Landfill gas is responsible for 2 to 4 percent of total greenhouse gas emissions worldwide (Markgraf 2016). But the release of the methane can be suppressed, for example by covering the waste with a thin layer of soil or plastic each night. If a methane capture system is installed, the landfill gas can be captured and converted into electricity. Installing the system requires upfront spending, and the payoffs come over time through the income stream from the electricity. The profitability of such an investment will depend on whether the borrower has access to sufficient capital at a low enough interest rate. One projection estimates that the adoption of landfill methane capture in 70 percent of the world's landfills would have an initial cost of around $35 billion and then would subsequently avert 2.2 billion metric tons of CO_2-equivalent by 2050 (Hawken 2017), implying a cost of roughly $22 per tCO2e, after accounting for time discounting. (For comparison, the world emitted 55 billion tCO2e in 2021, and the US Environmental Protection Agency estimates that the economic and social benefits of mitigation are $51 per tCO2e averted [US Environmental Protection Agency 2022]). Almost all of the untapped opportunity to fix highly-emitting landfills is in low- and middle-income countries (Maasakkers et al. 2022).

Another example of this phenomenon, at the household level, is adoption of energy-efficient cooking stoves. Berkouwer and Dean (2022) find that many households in Kenya do not adopt more energy-efficient stoves that reduce emissions and save the user money from fuel costs because they are credit-constrained. For a new energy-efficient stove that reduced charcoal use by 39 percent and saved households $237 from reduced charcoal use over its two-year lifespan, the average willingness to pay was only $12, which partly stems from a low ability to pay. Being offered a loan doubled the willingness to pay. Note that credit is not the only barrier: surely, there are other factors like inertia. This pattern of not adopting more energy-efficient options has been called the "energy paradox" and is seen in high-income countries, too (Gerarden, Newell, and Stavins 2017).

Weak Regulatory Capacity

While estimates often suggest methane capture is a negative cost opportunity, adoption in high-income countries has often come about only when required by regulations. The fact that landfill operators were not adopting these technologies out of self-interest suggests several possibilities. For example, perhaps there is a lack of arbitrageurs who are willing to try to bridge the gap between landfill companies and electricity grid companies. Another possibility is that landfill methane capture is not actually a negative cost opportunity (or at least not in all circumstances), but if environmental benefits are taken into account, it could still be a positive step for social welfare.

Lack of regulatory infrastructure might also explain why some truly negative-cost options are not being adopted. An opportunity might be negative-cost when all parties' costs and benefits are included, yet stymied by agency problems. Agency problems arise if the individual who makes the decision to invest and incurs the upfront costs is not the one who enjoys the cost savings that later accrue. For example, high-quality insulation in a home would save on heating and cooling bills, but builders might underinvest if they do not internalize the interests of the homeowner. Similarly, landlords will not have an incentive to invest in weatherization of a rental unit if their tenants pay the energy bills. In theory, if these investments could be verified, the builder or landlord could recoup the investment through a higher selling price or higher monthly rent. But if the investments cannot be verified easily, or if homebuyers and renters are inattentive to them when they make decisions, regulation can be helpful in achieving the socially efficient outcome. In such cases, regulation essentially requires one agent to behave in the way that is aligned with the other agent's financial interests.

In the case of landfills, landfill operators could be required to use a covering over the waste or to install landfill gas capture systems. For this to be a viable way to change behavior, the government needs to have the capacity to enact the regulation, monitor compliance, and punish those not in compliance. However, regulatory capacity is often limited in low- and middle-income countries (Besley and Persson 2009). In particular, when the sector that needs to be regulated is diffuse, with many different actors to monitor, it might be a major challenge. But for concentrated sources of emissions, it seems feasible for many middle-income countries to take at least some steps to strengthen regulatory capacity. For example, Duflo et al. (2013) evaluated a successful regulatory reform in Gujarat, India, that disallowed industrial plants from choosing their own pollution inspectors, introduced auditing of the inspectors' work to check if their reports were accurate, and paid inspectors based on their accuracy. The intervention reduced by 80 percent the likelihood that inspectors falsely reported that noncompliant plants were compliant.

Thus, one investment opportunity is to fund the regulatory infrastructure that would enable more low-cost mitigation. Funding is also useful for lessening the political resistance to regulation. For example, funds could be used to help defray firms' costs to upgrade their plants to be in compliance, so that their profits take less of a hit. Compensating firms would raise the project's costs, but in some cases, these projects would still be bargains, when taking a global perspective.

Reason #2: Lower Opportunity Costs of Inputs in Low- and Middle-Income Countries

The cost of mitigation in different countries will depend on the opportunity cost of the inputs used to achieve the mitigation. Where production of the abatement technology is intensive in inputs that are either sufficiently mobile that prices are similar across countries (for example, minerals) or particularly scarce and higher-priced in low- and middle-income countries (like capital and highly-skilled labor), there may be no advantage in locating the mitigation in low- and middle-income countries. For example, forms of carbon capture from the air that are intensive in capital and highly-skilled labor may be best located in high-income countries, where these factors are more abundant (Wilberforce et al. 2021).

However, many important mitigation opportunities have a high intensity of immobile factors which have lower absolute prices in low- and middle-income countries, because the opportunity cost of their use is lower. In particular, many mitigation investment opportunities require substantial inputs of land and unskilled labor, which can make the same type of mitigation activity less expensive in low- and middle-income countries than in high-income countries.

For example, forest preservation and reforestation are intensive in land and unskilled labor. Many high-income countries have highly ambitious tree-planting plans as part of their net zero and biodiversity commitments: the UK Chancellor announced plans to plant 30,000 new hectares of forest every year, while the European Union has announced plans to plant three billion trees by 2030. Both the economic and financial cost of reserving land for forest preservation or reforestation depends on the value of the alternative use to which this land would be put; for example, given that the main alternative use for forest land is agriculture, the opportunity cost depends on agricultural productivity per hectare in different countries.[2]

We consider three alternative approaches to estimating the economic and financial cost of forest preservation/reforestation by country. None of the approaches are without problems, but all suggest forest preservation/reforestation are between three and ten times more expensive in high-income countries than in low- and middle-income countries. Specifically, we compare cereal yields per acre, agricultural land rental values, and direct costs of forest preservation programs in different locations.

Cereal yields are a direct measure of the lost output if a hectare of land is moved from agricultural production to forest. As shown in Table 2, cereal yields per acre vary dramatically and are nearly ten times higher in New Zealand than in the Democratic Republic of Congo. However, this difference overestimates the difference in opportunity costs of land, at least to some extent, because cereal production in New

[2] The economic cost is the output forgone from using the land for forest, while the financial cost is the monetary price that would have to be paid to preserve the land for forest which in turn reflects the return to alternative uses. In a perfectly competitive market, the two would be equal.

Table 2
Cereal Yields in Selected Countries and Regions with Forests (2020)

	Tons per hectare
Low and middle income countries with over 90 million hectares of forest	
Brazil	4.479
Indonesia	5.351
China	6.319
Democratic Republic of Congo	0.877
Low cereal yields but high forest coverage	
Liberia	1.065
Tanzania	1.651
Congo	0.883
Gabon	1.589
High income, high cereal yield	
United States	8.268
France	7.171
Germany	6.998
United Kingdom	6.967
New Zealand	8.728

Source: Ritchie, Roser, and Rosado (2020c).
Note: Cereal yield is one proxy measure of the opportunity cost of allocating land to forest. Cereals include wheat, rice, maize, barley, oats, rye, millet, sorghum, buckwheat, and mixed grains.

Zealand uses many more inputs (including fertilizer, capital, and high skilled labor) that also have opportunity costs.

An alternative measure of the opportunity cost of land is land rental value, which abstracts from differences in agricultural inputs between countries. In many low-income countries, land rental markets are distorted by challenges in land registration and collective land ownership, but they still tend to function better than land sales markets. Not all land is suitable for all mitigation purposes: desert land may be cheap but not suitable for forest preservation. Thus, we compare rental rates for currently productive agricultural land to lessen this concern, though we cannot eliminate it.

Abay et al. (2021) use data from the Living Standards Measurement Survey to estimate rental prices of agricultural land in selected sub-Saharan African countries. Mean prices (updated to 2020 US dollars) range from \$56 per hectare per year in Malawi to \$170 in Ethiopia, as shown in Table 3. In contrast, farm rents in the United Kingdom are \$237 per hectare—that is, it is more than four times as expensive to rent agricultural land in the United Kingdom as in Malawi. This may well be an imprecise measure of the opportunity cost of using land in sub-Saharan Africa for reforestation/preserving forest, because Africa's land rental markets are not very developed. We therefore turn to a third way of estimating the relative cost of reforestation/preserving forest in different countries that comes closest to the full cost of preservation, but for which less data is available.

Table 3
Rental Prices of Agricultural Land

	Rental price per hectare (in 2020 USD)	Ratio of England's to other country's rental price
Ethiopia	$170.46	1.4
Malawi	$55.81	4.3
Tanzania	$72.74	3.3
England	$237.24	1.0

Source: Abay et al. (2021); Department of Environment, Food, and Rural Affairs (2023); International Monetary Fund (2023).
Note: This table compares the rental price of land, measured in US dollars per hectare, across select countries. Ethiopia's (value from 2013), Malawi's (value from 2013), and Tanzania's (value from 2015) prices are from Abay et al. (2021). England's price is from the Department of Environment, Food, and Rural Affairs (2021). Prices are converted to 2020 US dollars using the US inflation rate series from the International Monetary Fund (2023).

The opportunity cost of forest preservation/reforestation is not just potential alternative uses for land, because it also requires labor to manage forests. Some of the labor is relatively low-skilled; other parts are higher-skilled. The most comprehensive comparator of the economic and financial costs of forest preservation/reforestation is therefore the direct cost of programs seeking to preserve or restore forests in different countries. Given the distortions in the market for land sales and the fact that creating forest reserves would often require displacing communities, a common approach to land conservation is to pay people to undertake conservation on their privately-owned land. This approach also mitigates concerns that current users of the land may not have sufficient say in decisions on land sales or rentals, ensuring that they are getting sufficient compensation because they choose voluntarily to enter an agreement to preserve forests. To attract participants, the payments need to cover the person's costs to conserve, which includes lost income from the land and compensation for any labor. Jayachandran et al. (2017) found that offering households in Uganda just $28 per hectare a year not to cut down forest was successful in reducing deforestation by 50 percent, and with this program CO_2 emissions were avoided at a cost of $4 to $20 per metric ton (depending on assumptions). The similar Conservation Reserve Program, run by the US government, had a cost per metric ton of CO_2 avoided that was over ten times as high as the Ugandan program (Claassen, Cattaneo, and Johansson 2008; Jayachandran et al. 2017).

Although we have focused our discussion here on mitigation investments to protect or restore forests, these of course are not the only mitigation investments where the differential opportunity costs of land and labor are important determinants of mitigation costs. Solar farms and the water storage lakes associated with hydroelectric dams also have large land footprints; for example, Lovering et al. (2022)

show that ground mounted solar requires five times the land area per unit of energy produced as natural gas. While solar and hydro have important location restrictions (they need to be near large pools of electricity demand, which would otherwise be serviced by electricity generated by fossil fuels if they are to be cost-effective mitigation investments), these conditions are met in many low- and middle-income countries. Another example is enhanced rock weathering, a carbon capture technique that entails spreading certain kinds of crushed rock that absorb atmospheric carbon on fields. The technique requires land, labor, and access to mining residue, making middle-income countries well-positioned to implement it.

Reason #3: Build Green versus Retrofit Green

Most of the infrastructural growth in the world will be in low- and middle-income countries in the coming decades. Infrastructure in these countries is at present relatively underdeveloped, and population growth and urbanization will be faster in many of these countries. Indeed, three-quarters of the world's urban infrastructure that will exist in 30 years is yet to be built (Dasgupta 2018). India, China, and Nigeria will alone account for about 35 percent of project urban growth by 2050 (UN DESA 2018).

It is substantially cheaper to "build green" in low- and middle-income countries than to "retrofit green" in high-income countries. The central reason is that the choice set is larger at the planning stage than after construction has occurred. At the planning stage, a builder has many options for achieving an energy-efficient goal that differ in cost, one of which is to ignore the goal temporarily and then plan to retrofit later. But when construction is planned with energy conservation goals in mind, one of the builder's other options is very likely to be cheaper than retrofitting. After all, retrofitting involves an extra step of disassembly. Removing existing windows from a building and then replacing them with double-paned ones requires an extra step over installing the double-paned windows from the get-go. Indeed, retrofitting can sometimes even be more expensive than the entire cost of new construction, not just the extra costs to build green, because the decision-making about how to deconstruct and reconstruct often relies on specialist knowledge to understand structural considerations (Re Cecconi, Khodabakhshian, and Rampini 2022).

One major opportunity is to reduce the need for air conditioning through so-called "passive cooling." Natural ventilation, in addition to choices about space configuration and building materials, can help maintain lower temperatures indoors when it is hot outdoors. This need would be great even without rising global temperatures, simply because demand for air conditioning increases with income. By 2050, energy demand from air conditioning is projected to be over ten times the level in 2000, driven mostly by low- and middle-income countries (Isaac and van Vuuren 2009). Another energy-saving opportunity that is easier to capitalize on at the initial building stage is to create a district cooling system that provides air conditioning to several interconnected buildings using a centralized cooling plant.

Building green versus retrofitting applies not just to buildings, but also to other infrastructure such as transportation systems. The example of transportation infrastructure makes salient another cost of retrofitting, which is the disruption to people using the existing infrastructure. Constructing new public transportation in a crowded city often requires closures or causes slowdowns, in addition to displacement of business establishments and people. In contrast, in a nascent city, that infrastructure can be built with less disruption to existing patterns of life. One example of transportation systems being constructed in many cities is bus rapid transport, which involves a dedicated lane for buses so that they can travel rapidly even when regular lanes of traffic are jammed (Carrigan et al. 2013). Bus rapid transport has been especially popular in low- and middle-income countries. It reduces greenhouse gas emissions by displacing private vehicles and low-occupancy public transportation (like "matatus," which are privately owned mini-buses used as shared taxis), that may use older and more polluting fuels and vehicle technologies.

Reason #4: General Equilibrium Effects and the Benefits of Targeting

The integrated nature of the world economy means that action to mitigate carbon emissions in one country is likely to be partially muted by offsetting impacts in other countries. For example, if the extent of deforestation to produce palm oil in Indonesia or beef in Brazil declines, then the global price of these commodities will be higher than they otherwise would have been, encouraging others to cut down forest to produce more palm oil and beef. These general equilibrium effects mean that the impact of mitigation may be less than a partial equilibrium estimate would suggest.

But even if general equilibrium effects mean a reduction in a harmful activity in one location is offset by an increase in the same activity elsewhere (and evidence suggests it is not nearly so high), there can still be large gains from shifting an activity from a high- to low-carbon-intensity location.

Palm oil production offers a good illustration: here we draw on Hsiao's (2022) detailed work on the carbon impact of palm oil, which represents 5 percent of all CO_2 emissions from 1990 to 2016. About five-sixths of the world's palm oil is produced in Malaysia and Indonesia. Some of the CO_2 emissions from palm oil production come from the destruction of the forests to make way for the palm trees. However, almost 90 percent of the emissions arise from the destruction of peat, which forms the first layer of soil under some parts of the forests. Refraining from palm oil production in the parts of Indonesia and Malaysia where the activity destroys peat would likely lead to some displaced deforestation: deforestation will occur elsewhere to meet the global demand for vegetable oil. But as Hsiao (2022) shows, production of the main alternative vegetable oils does not involve the destruction of peat; for example, growing more soybeans in Brazil. Thus, protection of the forest in Indonesia and Malaysia that has peat would achieve 90 percent of its

CO_2 benefit—even if it led to exactly the same amount of forest being cut down in another place—as long as that alternative forest did not have peat substrate.

Taking general equilibrium effects and displacement into account when prioritizing mitigation efforts, in addition to targeting mitigation efforts to geographic areas with high carbon costs compared to when the same activity is done elsewhere, requires global thinking. Because (relatively) climate-conscious high-income countries are more likely to have regulations against further destruction of areas with the highest carbon equivalent impact, and because high-income countries only represent roughly one-quarter of the Earth's land surface, it is unlikely that the big gains from this type of targeting strategy will occur within the borders of high-income countries.

The Economic Counterargument

Several economic factors push in the opposite direction from the reasons we have highlighted so far and suggest carbon mitigation in high-income countries can be potentially more cost-effective.

First, carbon emissions are higher per person in high-income countries, while low-income countries emit hardly any carbon at all. Some efforts to mitigate carbon emissions are most cost-effective when emission levels are high. For example, a US office building is typically kept at a lower temperature in summer than a comparable building in a poorer country. Insulating the US building has more payoff in reduced energy use, because the cooling use is more intense. The same reasoning applies more broadly to legal and regulatory change; for example, the reduction in carbon emissions that results from passing regulation to limit emissions from cars will be higher in countries with higher car ownership, all else equal.

Monitoring and enforcement, which are needed in a range of mitigation activities from payments for conservation to regulation, are more challenging in low- and middle-income countries. However, it is wise to be cautious about drawing strong conclusions on this topic. People's perceptions of monitoring and enforcement of carbon mitigation actions in low- and middle-income countries are often formed from their knowledge of voluntary carbon "offset markets" (Bushnell 2010). These offset markets allow individuals in high-income countries to meet their carbon mitigation targets by, say, financing tree-planting in some low-income country. The carbon reductions from such offsets are indeed often dubious. But these offset markets are typically monitored by nongovernmental organizations with varying degrees of competence and limited checks on their over-claiming carbon benefits. Such voluntary offset markets may not say much about the ability of governments in these countries to carry out monitoring and regulation of carbon mitigation efforts. However, one implication of this concern is that the returns to investing in high-quality, easy-to-implement monitoring of carbon mitigation action and outcomes may be high, a theme to which we return below.

Despite these counterarguments, we believe that for many important types of carbon mitigation, the costs are likely to be substantially lower in middle-income countries that already have relatively high and growing carbon emissions than in high-income countries, even when enforcement and lower carbon intensity are considered. This is particularly true for countries like China, India, Indonesia, and Pakistan, where climate warming emissions are on the rise, relatively cheap opportunities to reduce emissions have not yet been tapped, and policies and programs can be introduced at scale. While enforcement may not be as good as in high-income countries, the lower costs are likely to more than compensate for this.

For low-income countries, the calculus for costs of mitigation is somewhat different, because the quite low levels of energy use per capita in these countries can make it cost-inefficient to switch to renewables if there are fixed costs of switching. For example, when rural households in Kenya are subsidized to connect to the electricity grid they do move away from fuels like kerosene (which produces black carbon, a particularly damaging warming gas). However, the high fixed costs in linking households to the electricity grid and very small quantities of energy used per household (both before and after) made the approach overall cost-inefficient (Lee, Miguel, and Wolfram 2020). The small scale of landholdings in many low-income countries can also drive up the costs of getting access to sufficient land to undertake carbon mitigation investments at scale. However, land availability varies substantially by country, and as the Uganda payments-for-conservation example above shows, it is possible to introduce programs that involve large numbers of small farmers to achieve meaningful mitigation impacts.

The Ethical Counterargument

Consider two premises: (1) high-income countries are willing to commit substantial resources to carbon mitigation; and (2) our arguments that it is often more cost-effective to do carbon mitigation in low-and middle-income countries hold weight. It follows that if high-income countries were to transfer some of the resources they are already willing to spend to finance carbon mitigation in low- and middle-income countries, greater and faster progress on the important goal of carbon mitigation could occur. Such a scenario offers potentially large gains from trade, and thus scope for all nations to enjoy some of those gains.

A natural question then becomes how to assure that such policies are enacted with the consent of host countries, and certainly not carried out in a way that impoverishes or otherwise harms them. There are ethical arguments against this type of global marketplace for mitigation. Some of these critiques are based on the premise that there should not be market transactions for certain goods or services, or that such exchanges cannot be truly voluntary. We briefly discuss three ethical arguments against international trade in emissions reductions.

One objection to trading in environmental protection is that no one should be detached from the world's problems, sacrificing only with their money and not their time or convenience or physical comfort. Goodin (1994) likens paying to offset environmental damage to the medieval practice of purchasing “indulgences” from the Catholic church to have one’s sins forgiven. Those with money have license to do wrong things (like carbon emissions) and then absolve themselves of blame through money. Sandel (2012) makes a related, consequentialist argument that trading in environmental protection may erode people’s sense of caring about the environment, and in that sense prove ultimately counterproductive.

This concern is multifaceted. In conventional pollution control, firms in polluting industries buy permits from the government to operate, and few would argue that such a policy is unethical because it is nothing more than allowing them to spend money to absolve themselves of blame. Likewise, firms with deeper pockets can afford the pollution-abatement equipment required to meet regulatory standards. It is not clear why using one’s financial resources to reduce pollution domestically is ethically acceptable, but then becomes unacceptable if it involves a payment to reduce pollution in another country. Moreover, we certainly do not envision that efforts to mitigate carbon pollution in high-income countries would be eliminated, only that some of the resources would have greater effect if spent in low- and middle-income countries.

In consequentialist terms, attempting to create a regime in which each person takes on the same personal sacrifice in nonmonetary terms would come at a high cost. Given limited resources to spend on climate mitigation, such a rule would probably have the practical effect that less mitigation would take place. As discussed earlier, low- and middle-income countries are expected to be hardest hit by climate change—both by the temperature and weather effects, and also in their lower level of resources to address these consequences—and so less success in reducing CO_2 levels would be devastating for the global poor. In addition, by passing up on these opportunities, the global poor would lose their share of the gains from trade alluded to above.

A second concern about environmental markets questions how voluntary they are. For example, Satz (2010) questions whether exchange can be truly voluntary when one party is vulnerable or desperate, using a motivating example of a poor country that needs money for basic services so badly that it agrees to house toxic waste (with potential long-term consequences for health and productivity) in exchange for immediate payments from a richer country. Arguably, the injustice here is the poor country’s lack of good options, not the exchange. Preventing low-income countries from making choices they believe are welfare-improving for them risks making them worse off. Moreover, this specific concern seems less applicable for the case of climate change mitigation, as most mitigation projects have positive co-benefits for low- and middle-income countries, such as reduced conventional air pollutant levels or perhaps technology transfer. These incidental benefits represent some of the gains from trade enjoyed by low- and middle-income countries—how they benefit from being the site of mitigation activity.

A third concern is that even if the exchange is voluntary for the party engaging in it—say, the national or local government—it might not be voluntary and beneficial for the many individuals who are affected. Governments might be corrupt. Politicians might personally benefit, while their constituents do not. Costs imposed on the most powerless in society, like indigenous groups, may be ignored. We view this argument as especially pertinent and believe there is an obligation for the high-income countries to consider and discuss openly the distributional consequences of a mitigation project within low- and middle-income countries. While countries have sovereignty over their people, a high-income country should not lean on sovereignty to fund efforts that knowingly exacerbate poverty or lead to other harms. Nor, however, should high-income countries be paternalistic and assume that people poorer than themselves cannot make rational choices. A practical way forward is to prioritize paying for mitigation in democracies with functioning land rights and other legal rights, where compensation mechanisms are more likely to work and exploitative behavior, if it occurs, is likely to surface and become apparent.

Some Policy Implications

In this section, we ask what policies should be pursued to capture the carbon reduction benefits available from investing a larger proportion of carbon mitigation financing in low- and middle-income countries. We discuss to what extent carbon pricing—specifically, a carbon border adjustment tax—would help encourage the switch in the location of mitigation activity (the answer is only partly). Second, we look at the practical steps that would be needed to implement the vision we have laid out, in which much more mitigation activity takes place in low- and middle-income countries.

Would a Carbon Border Adjustment Tax Encourage Efficient Allocation of Mitigation Activity?

Carbon pricing is often the economist's first-best solution to climate change challenges, and imposing a tax on greenhouse gas emissions (and credit for carbon storage activity) in all countries would help address some of the missed low-cost opportunities raised in this article. However, a universal carbon tax is a long way from being a political reality, and the more realistic alternative, already under development in the European Union, is a carbon border adjustment tax. Under a carbon border adjustment tax, imports are taxed based on the carbon content of the product and the carbon pricing in the exporting country. The adjustment seeks to equilibrate the carbon price of domestically produced and imported goods, creating an incentive for mitigating actions in the exporting country. The tax discourages rich countries from specializing in greener industries and just outsourcing dirtier production to other, usually poorer countries.

There is a large economic literature on carbon border adjustment taxes, and we do not attempt to summarize it here (for a starting point, the interested reader might begin with Fontagné and Schubert 2023). Instead, we point out that even if carbon border adjustment taxes were implemented by all high-income countries, this would not achieve the gains in reduced CO_2-equivalent emissions from thinking globally about emissions opportunities we discuss in this paper.

First, carbon border adjustment taxes require substantial information on the carbon intensity of the product, and thus are likely to only be imposed (at least initially) on a few products with high carbon intensity, such as steel. Second, carbon border adjustment taxes will only encourage countries to engage in mitigation on production of goods destined for export to countries imposing such a tax. They create no incentives for mitigating action on products produced for domestic consumption or for export to countries that do not impose such a tax. Third, carbon border adjustment taxes provide no incentive for low- and middle-income countries to undertake mitigation actions that are divorced from production, such as forest protection, capturing methane from waste, or enhanced rock weathering.

Finally, while carbon border adjustment taxes are likely to produce positive carbon reduction gains, they could have negative distributional impacts—effectively making low- and middle-income countries pay for mitigating carbon. In comparison, if high-income countries spend their mitigation funds wherever in the world achieves the highest impact and pay the full costs including appropriate compensation for local land and labor, such negative distributional impacts are far less likely to occur.

The Skeleton for a Workable System of International Carbon Mitigation Payments

Designing and enacting a workable system in which high-income countries redirect some of the resources they are allocating to climate change mitigation in their own countries to achieve a greater level of emissions in low- and middle-income countries is a substantial task. Here, we suggest four principles that could guide the design of such a policy.

First, the "nationally determined contributions" to carbon mitigation by each country must take into account mitigation efforts outside a country's borders. The current international climate agreements incorporate this principle, but much more work is needed to operationalize and encourage it. Countries have made commitments about how much they will reduce carbon emissions as part of the Paris Agreement and subsequent United Nations Conferences of the Parties. However, such commitments are primarily framed around emissions within the country's territory only. Thus, the incentive—really a distortion—is for countries to focus on carbon mitigation at home.

Article 6 of the Paris Agreement sketches out the possibility of cross-border trade in mitigation, either bilaterally (Article 6.2) or through a yet-to-be-established centralized marketplace (Article 6.4). The provisions for bilateral arrangements implicitly focus on high-income countries, while the centralized marketplace, through which one party would finance and receive emissions credit for another party's mitigation project, has yet to be established (Fattouh and Maino 2022).[3]

[3] For example, the bilateral trade would be between countries that have set absolute mass-based targets for their emissions reductions relative to a reference year, most of whom are high-income.

Making the ideas sketched out in Article 6 a reality, with a focus on unlocking the currently-underfunded opportunities in low- and middle-income countries, should be a priority for international climate policymakers.

Second, verification and monitoring of mitigation efforts need to be improved. Abatement projects purchased through the international marketplace need to truly reduce emissions to deliver on the promise of more mitigation for less money. The private-sector offset market mentioned earlier was meant to provide a transparent marketplace where different mitigation options around the world could be supplied, and buyers could invest in the most cost-effective ones wherever they were located. However, the market has limited credibility because of concern that the credits overstate the amount of mitigation genuinely generated by the projects. The Clean Development Mechanism, which was established under the Kyoto Protocol to enable high-income countries to invest in mitigation in low- and middle-income countries, suffered the same problem.

The key challenge is that credits are determined by comparing the actual carbon output with what would have happened otherwise. Both actual and counterfactual emissions are hard to estimate. This challenge is not specific to projects in low- and middle-income countries; it applies to any scheme that offers credit for investing in a mitigation project. But overcoming it is essential so that the marketplace is not giving high-income countries emissions credits for projects that would have happened anyway. A trustworthy intermediary that uses a rigorous standard when defining the counterfactual is a first step.

In addition, investment and innovation are needed to provide more objective, credible, and cheaper ways to monitor mitigation in low- and middle-income countries. For example, improved algorithms that use satellite data to construct more precise measures of the amount of carbon embodied in tree cover or other carbon-mitigating farming practices would be valuable. As development of these technologies are a global public good, they will be undersupplied by the private sector. Offering prizes for algorithms that can achieve these goals, which could then be made public, might be an efficient way to stimulate innovation in this area. While hard to operationalize, countries that invest in these innovations (or any other research and development that enables more cost-effective and larger-scale mitigation) ideally would receive credit for emissions reductions.[4]

It is also worth encouraging bilateral arrangements between high-income and low- or middle-income countries. While there are advantages of a trusted intermediary certifying projects, in bilateral arrangements, countries would perhaps face more reputational damage if they claim credits for projects that do not deliver on

[4] Investments in research and development to reduce carbon emissions can have high returns and, in general, are undersupplied both by the private sector and by governments seeking to achieve only their own climate targets. Ideally, the appropriate mitigation credit for a research and development investment would be calculated on the expected value of reduced emissions, so that funders do bear the risk that their particular R&D effort does not succeed, and do not avoid risky investments.

actual emissions reductions or pursue projects that are exploitative of the local population.

Third, the anticipated local co-benefits of mitigation projects should be laid out explicitly, and then measured to the extent possible, to ensure that the low- and middle-income countries receive their fair share of the gains from trade. A political challenge to a robust international market for mitigation is that low- and middle-income countries may object to giving all the mitigation credit of a project funded by others, but within their borders, to the funding country. After all, they incur costs, perhaps to implement regulation or provide land. From an economist's perspective, the focus should be on how the gains from trade are shared rather than more narrowly on how mitigation credit is shared. The funding country could make financial transfers to the host country. There are also often incidental co-benefits, like reduced conventional pollution. Explicit accounting of the benefits for the low- or middle-income country, in whatever form they take, is important for ensuring that such arrangements are mutually beneficial.

Because such cross-border agreements would be voluntary, the choice of a low- or middle-income country to participate means that it expects to be made better off by participating. But both parties being better off does not pin down how the gains from trade are split between them. Here, internationally agreed-upon guidelines that ensure an equitable split would be valuable. The potential surplus from cross-country mitigation agreements is large, so it should be possible to make participating in this type of exchange attractive for both funding and host countries. Importantly, ensuring that the low- and middle-income participant receives a fair share of the surplus is different from the current Article 6 approach of requiring funders to contribute to a general adaptation fund whenever they fund mitigation projects in low- and middle-income countries. However well-intentioned, this "share of proceeds for adaptation" provision essentially taxes—and thus discourages—international trade in mitigation projects. Directing those proceeds specifically to the low- or middle-income country hosting the project would similarly shift financial resources from rich to poor countries, but without stifling international cooperation that could help the world achieve lower emissions.

Fourth, mitigation in low- and middle-income countries should not be treated as development aid. Currently, when high-income countries do fund mitigation projects in low- and middle-income countries, it is often thought of and counted as foreign aid. The justification is that there are co-benefits that accrue to the country where the project operates; for example, switching from coal to solar electricity reduces local particulates which otherwise would damage local health.[5] But as long as mitigation in low- and middle-income countries is seen as aid, and not as a central part of high-income countries own effort to reduce emissions, it will always be small and fail to reflect the potentially large gains set out in this paper. Moreover, diverting aid

[5] On the one hand, local co-benefits are a reason to invest at home, to help one's own citizens. On the other hand, these co-benefits mean that investment in mitigation in low- and middle-income countries has a redistributive benefit.

budgets to mitigation risks reduces actual development assistance. The amount spent on a mitigation project is not a measure of the benefit to the low- or middle-income country. Only the increase in well-being in the low- or middle-income country, in the form of local environmental co-benefits or surplus from job creation, for example, constitute foreign aid to that country. This local surplus will often be modest relative to the total project budget, because many of the project costs are to offset the host country's opportunity costs of participating in mitigation projects or to purchase inputs produced elsewhere. Most of the spending should be considered an investment in the global public good of climate change mitigation, not aid.

These questions of how to quantify (and who gets credit for) mitigation outside a country's borders, and whether it should be considered aid or not, may seem arcane, but they are the key to unlocking additional emissions reductions as high-income countries redirect some of their mitigation spending to the highest-return locations.

A Coda on Funding Adaptation

While our discussion has mainly focused on mitigation of carbon emissions, we close by touching on the role of high-income countries in funding adaptation to climate change in low- and middle-income countries (Fankhauser 2017). The economic issues differ here. Mitigation efforts are a global public good: lower emissions in one part of the world help the rest of the world. In contrast, most efforts to help people adapt to climate change have their effects locally. A levee helps the community in which it is built. Food aid sent to a drought-stricken area helps people in that community cope. Thus, funding adaptation, unlike mitigation, in low- and middle-income countries is not a way for high-income countries to achieve their abatement goals faster or less expensively. Instead, it is a way to help the world's poor; many types of adaptation are squarely in the category of development aid.

However, some efforts to improve adaptation can represent, if not truly global public goods, at least cross-national public goods that could help in low- and middle-income countries at large. Aid agencies of high-income countries have an important opportunity to invest in these areas.

One example is research and development of technologies that facilitate adaptation. We have already noted that investment in technological innovation is undersupplied by the market and that social returns from innovation have been estimated at twice the private returns (Aghion, Van Reenen, and Zingales 2013). The additional distortions in markets in low- and middle-income countries suggest that that innovation is even more undersupplied for their challenges (Kremer and Glennerster 2004). Investment in climate-resilient crops offers an example. A large body of literature documents high returns to investments in innovation in agriculture in low- and middle-income countries: one meta-analysis suggests an average return of 100 percent a year (Alston et al. 2000). Other work has found that countries with agriculture that is more distant (biologically) from that found in high-income countries have experienced the least innovation in the past, suggesting substantial further innovation is possible (Moscona and Sastry 2021). Finally, a randomized evaluation of the impact of introducing flood-tolerant rice (developed through the public research centers) showed a 10 percent increase in rice yields (Dar et al. 2013).

Social science research on behavior change could offer high returns, as well. Change is hard for people, and climate change will require people to change their habits and choices. Understanding how to encourage adoption of (say) the new drought-resistant seeds is essential. An example of how social science research can impact behavior change and technology adoption is the work on the drivers of health technology adoption that has arguably helped save millions of lives. As one example, the free mass distribution of antimosquito bed-nets has been estimated as reducing malaria deaths by four million in sub-Saharan Africa between 2000 and 2014 (Bhatt et al. 2015).

No amount of mitigation effort will avert climate change; climate change is already upon us. Thus, finding and pursuing the high-return opportunities for adaptation should also be a priority.

■ *We thank Jonathan Colmer, Rebecca Dizon-Ross, Elspeth Kirkman, and Rob Stavins for helpful comments and discussion; Erik Hurst, Nina Pavcnik, Timothy Taylor, and Heidi Williams for valuable editorial guidance; and Meghana Mungikar and Juliana Sanchez Ariza for excellent research assistance.*

References

Abay, Kibrom A., Guush Berhane, John Hoddinott, and Kibrom Tafere. 2021. "Assessing Response Fatigue in Phone Surveys: Experimental Evidence on Dietary Diversity in Ethiopia." International Food Policy Research Institute Discussion Paper 02017.

Aghion, Philippe, John Van Reenen, and Luigi Zingales. 2013. "Innovation and Institutional Ownership." *American Economic Review* 103 (1): 277–304.

Alston, Julian M., Connie Chan-Kang, Michele C. Marra, Philip G. Pardey, and T. J. Wyatt. 2000. "A Meta-analysis of Rates of Return to Agricultural R&D." International Food Policy Research Institute Research Report 113.

Armstrong McKay, David I., Arie Staal, Jesse F. Abrams, Ricarda Winkelmann, Boris Sakschewski, Sina Loriani, Ingo Fetzer, Sarah E. Cornell, Johan Rockström, and Timothy M. Lenton. 2022. "Exceeding 1.5°C Global Warming Could Trigger Multiple Climate Tipping Points." *Science* 377 (6611): eabn7950.

Berkouwer, Susanna B., and Joshua T. Dean. 2022. "Credit, Attention, and Externalities in the Adoption of Energy Efficient Technologies by Low-Income Households." *American Economic Review* 112 (10): 3291–330.

Besley, Timothy, and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99 (4): 1218–44.

Bhatt, S., D. J. Weiss, E. Cameron, D. Bisanzio, B. Mappin, U. Dalrymple, K. E. Battle et al. 2015. "The Effect of Malaria Control on *Plasmodium falciparum* in Africa between 2000 and 2015." *Nature* 526: 207–11.

Buchner, Barbara, Baysa Naran, Pedro Fernandes, Rajashree Padmanabhi, Paul Rosane, Matthew Solomon, Sean Stout et al. 2021. *Global Landscape of Climate Finance 2021.* San Francisco: Climate Policy Initiative.

Bushnell, James B. 2010. "The Economics of Carbon Offsets." NBER Working Paper 16305.

Carrigan, Aileen, Robin King, Juan Miguel Velásquez, Matthew Raifman, and Nicolae Duduta. 2013. *Social, Environmental and Economic Impacts of BRT Systems*. Washington, DC: World Resources Institute.

Claassen, Roger, Andrea Cattaneo, and Robert Johansson. 2008. "Cost-Effective Design of Agri-environmental Payment Programs: U.S. Experience in Theory and Practice." *Ecological Economics* 65 (4): 737–52.

Dar, Manzoor H., Alain de Janvry, Kyle Emerick, David Raitzer, and Elisabeth Sadoulet. 2013. "Flood-Tolerant Rice Reduces Yield Variability and Raises Expected Yield, Differentially Benefitting Socially Disadvantaged Groups." *Science Report* 3: 3315.

Dasgupta, Ani. 2018. "IPCC 1.5° Report: We Need to Build and Live Differently in Cities." *World Resources Institute*, October 31. https://www.wri.org/insights/ipcc-15deg-report-we-need-build-and-live-differently-cities.

Department for Environment, Food and Rural Affairs. 2021. "Farm Rents-England 2021." London: UK Department for Environment, Food and Rural Affairs. https://www.gov.uk/government/statistics/farm-rents/farm-rents-england-2020#:~:text=For%20those%20agreements%20with%20a,to%20%C2%A3190%20per%20hectare (accessed June 5, 2023).

Duflo, Esther, Michael Greenstone, Rohini Pande, and Nicholas Ryan. 2013. "Truth-Telling by Third-Party Auditors and the Response of Polluting Firms: Experimental Evidence from India." *Quarterly Journal of Economics* 128 (4): 1499–545.

Dupas, Pascaline. 2014. "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment." *Econometrica* 82 (1): 197–228.

Fankhauser, Sam. 2017. "Adaptation to Climate Change." *Annual Review of Resource Economics* 9: 209–30.

Fattouh, Bassam, and Andrea Maino. 2022. *Article 6 and Voluntary Carbon Markets*. Oxford, UK: Oxford Institute for Energy Studies.

Fontagné, Lionel, and Katheline Schubert. 2023. "The Economics of Border Carbon Adjustment: Rationale and Impacts of Compensating for Carbon at the Border." *Annual Review of Economics*. https://doi.org/10.1146/annurev-economics-082322-034040.

Gerarden, Todd D., Richard G. Newell, and Robert N. Stavins. 2017. "Assessing the Energy-Efficiency Gap." *Journal of Economic Literature* 55 (4): 1486–525.

Glennerster, Rachel, and Seema Jayachandran. 2023. "Replication data for: Think Globally, Act Globally: Opportunities to Mitigate Greenhouse Gas Emissions in Low- and Middle-Income Countries." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E192082V1.

Goodin, Robert E. 1994. "Selling Environmental Indulgences." *Kyklos* 47 (4): 573–96.

Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano J., M. Lagos, P. Norris, E. Ponarin & B. Puranen (eds.). 2022. "World Values Survey: Round Seven—Country-Pooled Datafile Version 5.0." Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat. doi:10.14281/18241.20.

Hawken, Paul. 2017. *Drawdown*: *The Most Comprehensive Plan Ever Proposed to Reverse Global Warming*. New York: Penguin.

Hoegh-Guldberg, O., D. Jacob, M. Taylor, T. Guillén Bolaños, M. Bindi, S. Brown, I. A. Camilloni et al. 2019. "The Human Imperative of Stabilizing Global Climate Change at 1.5°C." *Science* 365 (6459): eaaw6974.

Hsaio, Allan. 2022. "Coordination and Commitment in International Climate Action: Evidence from Palm Oil." Unpublished.

International Monetary Fund (IMF). 2023. "Inflation rates, end of period consumer prices." World Economic Outlook (WEO) database. https://www.imf.org/external/datamapper/datasets/WEO (accessed June 20, 2023).

International Monetary Fund (IMF). 2023. "GDP per capita, current prices. Purchasing power parity; international dollars per capita." World Economic Outlook (WEO) database. https://www.imf.org/external/datamapper/datasets/WEO (accessed June 23, 2023).

Isaac, Morna, and Detlef P. van Vuuren. 2009. "Modeling Global Residential Sector Energy Demand for Heating and Air Conditioning in the Context of Climate Change." *Energy Policy* 37 (2): 507–21.

Jayachandran, Seema, Joost De Laat, Eric F. Lambin, Charlotte Y. Stanton, Robin Audy, and Nancy E. Thomas. 2017. "Cash for Carbon: A Randomized Trial of Payments for Ecosystem Services to Reduce Deforestation." *Science* 357 (6348): 267–73.

Kremer, Michael, and Rachel Glennerster. 2004. *Strong Medicine: Creating Incentives for Pharmaceutical Research on Neglected Diseases.* Princeton, NJ: Princeton University Press.

Lee, Kenneth, Edward Miguel, and Catherine Wolfram. 2020. "Experimental Evidence on the Economics of Rural Electrification." *Journal of Political Economy* 128 (4): 1523–65.

Lovering, Jessica, Marian Swain, Linus Blomqvist, and Rebecca R. Hernandez. 2022. "Land-Use Intensity of Electricity Production and Tomorrow's Energy Landscape." *PLoS ONE* 17 (7): e0270155.

Maasakkers, Joannes D., Daniel J. Varon, Aldís Elfarsdóttir, Jason McKeever, Dylan Jervis, Gourav Mahapatra, Sudhanshu Pandey et al. 2022. "Using Satellites to Uncover Large Methane Emissions from Landfills." *Science Advances* 8 (32): eabn9683.

Markgraf, Claire. 2016. *Financing Landfill Gas Projects in Developing Countries.* Washington, DC: International Bank for Reconstruction and Development/World Bank.

Moscona, Jacob, and Karthik Sastry. 2021. "Inappropriate Technology: Evidence from Global Agriculture." Unpublished.

Parry, Ian, Peter Dohlman, Cory Hillier, Martin Kaufman, Kyung Kwak, Florian Misch, James Roaf, and Christophe Waerzeggers. 2021. "Carbon Pricing: What Role for Border Carbon Adjustments?" IMF Staff Climate Note 2021/004.

Re Cecconi, Fulvio, Ania Khodabakhshian, and Luca Rampini. 2022. "Data-Driven Decision Support System for Building Stocks Energy Retrofit Policy." *Journal of Building Engineering* 54: 104633.

Ritchie, Hannah, Max Roser, and Pablo Rosado. 2020a. "CO_2 and Greenhouse Gas Emissions." Oxford: Our World in Data. https://ourworldindata.org/co2-emissions (accessed June 20, 2023).

Ritchie, Hannah, Max Roser, and Pablo Rosado. 2020b."Share of Electricity Production by Source, Upper-Middle-Income Countries." Oxford: Our World in Data. https://ourworldindata.org/grapher/share-elec-by-source?country=~Upper-middle-income+countries (accessed June 20, 2023).

Ritchie, Hannah, Max Roser, and Pablo Rosado. 2020c. "Crop Yields." Oxford: Our World in Data. https://ourworldindata.org/crop-yields (accessed June 20, 2023).

Sandel, Michael J. 2012. *What Money Can't Buy: The Moral Limits of Markets.* New York: Farrar, Straus and Giroux.

Satz, Debra. 2010. *Why Some Things Should Not Be for Sale: The Moral Limits of Markets.* Oxford, UK: Oxford University Press.

UN Climate Change. 2021. "COP26: The Glasgow Climate Pact." https://ukcop26.org/wp-content/uploads/2021/11/COP26-Presidency-Outcomes-The-Climate-Pact.pdf (accessed September 19, 2022).

UN Department of Economic and Social Affairs (DESA). 2018. "68% of the World Population Projected to Live in Urban Areas by 2050, Says UN." *UN DESA News*, May 16. https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html.

UN Environment Programme (UNEP). 2021. *Emissions Gap Report 2021: The Heat Is On–A World of Climate Promises Not Yet Delivered*. Nairobi, Kenya: UNEP.

US Environmental Protection Agency. 2015. "Overview of Greenhouse Gasses." https://www.epa.gov/ghgemissions/overview-greenhouse-gasses.

US Environmental Protection Agency. 2022. "Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances." https://www.epa.gov/system/files/documents/2022-11/epa_scghg_report_draft_0.pdf (accessed June 20, 2023).

Wilberforce, Tabbi, A. G. Olabi, Enas Taha Sayed, Khaled Elsaid, and Mohammad Ali Abdelkareem. 2021. "Progress in Carbon Capture Technologies." *Science of the Total Environment* 761: 143203.

Journal of Economic Perspectives—Volume 37, Number 3—Summer 2023—Pages 137–162

Carbon Border Adjustments, Climate Clubs, and Subsidy Races When Climate Policies Vary

Kimberly A. Clausing and Catherine Wolfram

Climate change is a global problem: carbon emitted anywhere contributes to atmospheric carbon levels everywhere, and policies that reduce carbon emissions benefit the entire world. However, climate change policy is usually adopted at the national (or subnational) level, and in a globally integrated economy, international trade can create important policy spillovers. For example, whenever a jurisdiction puts a price on carbon, it faces two concerns. First, its producers may face a competitive disadvantage, because other jurisdictions may not price carbon and might even subsidize energy. Second, the benefits of ambitious climate policy will be limited because the country will reap only a share of the gains; that is, any emissions reductions will benefit all jurisdictions, regardless of their policy stance. Both concerns may lead governments to adopt insufficiently ambitious climate policies.

As a starting point, a conceptual framework for climate change mitigation policies might usefully be thought of as occupying a two-by-two matrix, with policy ambition on one axis, and approach (taxes versus subsidies) on the other (see Table 1). For example, Canada (with a carbon price of about $50 per ton in 2023

■ *Kimberly Clausing is the Eric M. Zolt Chair in Tax Law and Policy, UCLA School of Law, Los Angeles, California. Catherine Wolfram is the William Barton Rogers Professor of Energy Economics, MIT Sloan School of Management, Cambridge, Massachusetts. When this paper was written, she was Visiting Raymond Plank Professor, Harvard Kennedy School, Cambridge, Massachusetts. Clausing and Wolfram are also Research Associates at the National Bureau of Economic Research, Cambridge, Massachusetts, and Clausing is a nonresident senior fellow at the Peterson Institute for International Economics. Their email addresses are clausing@law.ucla.edu and cwolfram@mit.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.3.137.

Table 1.
Matrix of Jurisdictions by Climate Mitigation Ambition and Approach

	High Ambition	*Low Ambition*
Cost Imposing	Country A	Country B
Cost Reducing	Country C	Country D

that is scheduled to rise) might be designated as primarily using a high ambition, tax-based approach, whereas Colombia (with a carbon price of $5 per ton) might be classified as emphasizing a low ambition tax-based approach. The United States could arguably be classified as emphasizing a low ambition cost-reducing approach up through 2021, but after passing the Inflation Reduction Act of 2022, it is arguably on track to use a high ambition, subsidy-based approach.[1] (Regulations can also be considered as an implicit tax, but the quantitative impact of such an implicit tax may be small and/or difficult to measure.)

In this essay, we begin by documenting the variation in climate change policies across countries by focusing on the rows of Table 1; that is, we describe policies that impose a cost on carbon and policies that seek to reduce the cost of shifting to less carbon-intensive processes. We discuss concerns raised by these policies, including effects on competition between nations that have adopted divergent approaches. For instance, cost-imposing jurisdictions may be concerned that their companies face a disadvantage when competing with companies based in jurisdictions that subsidize the transition to a low-carbon economy, particularly in industries that are traded both in a global market and carbon intensive, such as chemicals. These concerns can even lead to "subsidy races," a dynamic we discuss below.

In addition, we consider variation in climate change policies, focusing on the columns of Table 1. Countries undertaking ambitious policy action (either cost-imposing or cost-reducing) may be concerned that other countries will forgo strong climate policy measures, instead free-riding on the costly efforts of others. Countries that subsidize face fiscal costs, and cost-imposing countries that regulate or tax also face political economy costs from implementing cost-imposing policies. The political economy of bearing these costs may be affected by the number of other countries taking similar measures.

We then turn to two main proposals to address these policy spillovers. First, "carbon border adjustment mechanisms" seek to address competitiveness concerns by imposing costs on imports to reflect differences in climate policies across

[1] While we use the term "ambition," which sometimes has a normative connotation, to differentiate levels of climate policy moving forward, it is important to recognize that high-income countries have contributed a much larger share of the current stock of greenhouse gas emissions in the past.

countries. Second, "climate clubs" (as proposed by Nordhaus 2015) would have higher-ambition climate-policy "club" countries levy a broad tariff on lower-ambition "non-club" countries in order to inspire greater mitigation action. We describe the economic policy issues raised by each remedy, drawing on recent literature. By responding to the competitiveness concerns of domestic industry, as well as the fear that other countries will free ride on domestic efforts, enforcement mechanisms such as carbon border measures or climate clubs have the potential to enable more effective policies globally, but there are also important policy risks.

Throughout our discussion, we emphasize that beyond the static, immediate effects of these policies, the ways in which they drive the evolution of future policy may be even more important. Under what conditions would carbon border adjustments and climate clubs lead to a "race to the top" and encourage a globally harmonized approach to climate mitigation? Alternatively, might carbon border measures or climate clubs simply ignite trade disputes, eroding the gains from trade and undermining climate policy? In the absence of border measures, will the pressures of international competition and domestic politics in the presence of asymmetric approaches to carbon mitigation unravel the climate strategies of even the best-intentioned and most ambitious governments? Are there alternative ways to foster improved alignment of ambitious climate policy?

Carbon Pricing to Reduce Carbon Emissions

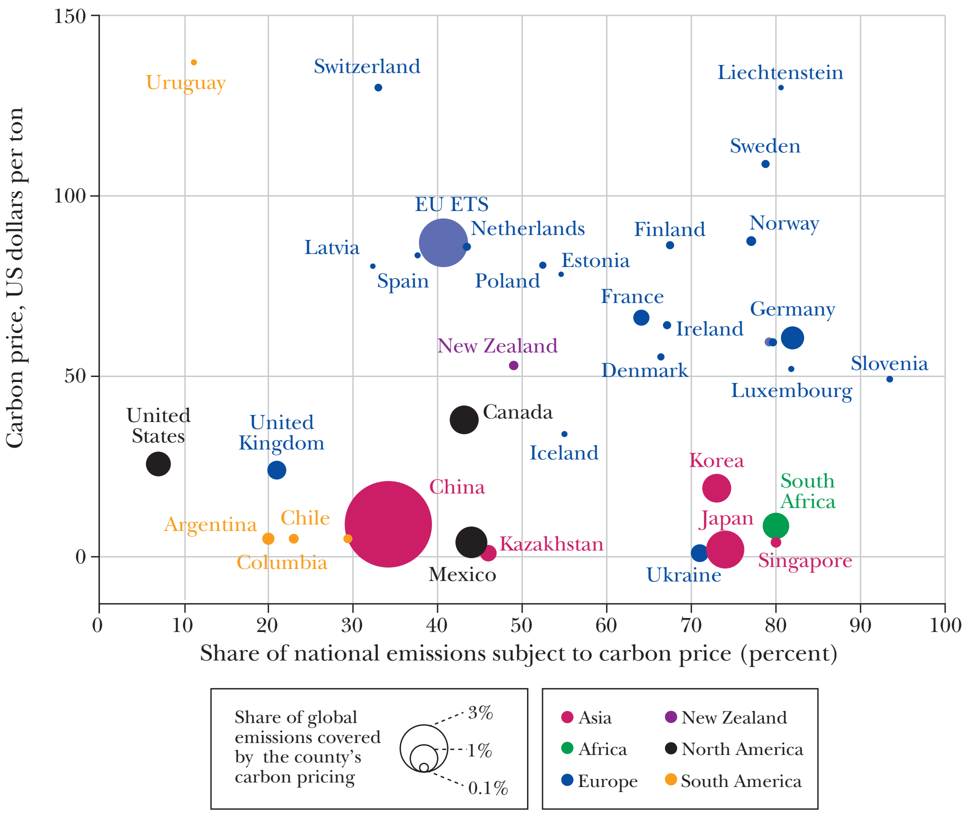
Carbon pricing efforts take multiple forms. Some jurisdictions price carbon directly and impose a carbon tax. Others price carbon by limiting emissions and then allowing trading of emissions permits in a "cap-and-trade" system: companies with excess emissions allowances may sell permits (and thus face an opportunity cost for emitting carbon), while those with deficit emissions allowances may buy them (and thus face a direct monetary cost for emitting carbon).

Cross-National Variation in Carbon Prices

Figure 1 shows the current state of carbon pricing efforts throughout the world, as compiled by the World Bank Carbon Pricing Dashboard, with adjustments by the authors to allocate the impacts of the supranational climate policies in the European Union to member countries. The horizontal axis shows the share of carbon emissions (or carbon-equivalent of other emitted greenhouse gases) in the country covered by carbon pricing. The vertical axis shows the price of carbon. The size of the circles is scaled to the share of global carbon emissions from that country. The colors of the circles refer to the continent where the emissions occurred. In countries with subnational policies or sector-specific policies, the graph reflects the weighted average carbon price. In the United States, for example, the weighted average price across jurisdictions with carbon pricing was about $25 per ton.

As of 2022, the World Bank Carbon Pricing Dashboard indicates that 70 jurisdictions—including 47 national jurisdictions as well as subnational

Figure 1
Carbon Pricing Coverage and Price Level by Jurisdiction



Source: World Bank Carbon Pricing Dashboard: https://carbonpricingdashboard.worldbank.org/ and authors' calculations.
Notes: Emissions refer to greenhouse gas emissions. Carbon price is per ton of CO^2-equivalent. Emissions covered by the EU Emissions Trading System are allocated to participating countries assuming an equal share of power and industrial emissions (69 percent) are covered in each country. We also represent the EU ETS as a whole. Note that these data do not reflect fossil fuel subsidies or taxes such as gasoline taxes.

jurisdictions—were subject to some form of explicit carbon pricing, covering 23 percent of all global greenhouse gas emissions. The policies in these jurisdictions do not just vary in terms of design; there is also substantial variation in the implied price of emissions. In April 2022, prices were over $80 per ton in the European Union, and even higher in some national European jurisdictions and Uruguay. Canada's price was $40 per ton in 2022, but it is scheduled to increase by $15 Canadian (about $11 in US dollars) per year between 2023 and 2030. California has the highest carbon price in the United States, at over $30 per ton in April 2022. Many subnational jurisdictions have modest carbon prices, including those in Japan and China, where carbon prices were under $15 per ton in mid-2022.

The share of emissions covered also varies. While 23 percent of greenhouse gas emissions worldwide are covered by some sort of pricing regime, that share is over 80 percent in Germany, 75 percent in Japan, 33 percent in China, and 78 percent in British Columbia. Within the nations belonging to the OECD, one-third of greenhouse gas emissions are covered by an explicit price, and the share rises to nearly 55 percent when the United States is excluded.[2] Multiplying the share of emissions covered by the carbon price times the price itself provides a sense of the economy-wide average carbon price, which varies from near-zero in many countries to over $50 per metric ton in Norway, Sweden, and Switzerland and over $100 per metric ton in tiny Liechtenstein.

International organizations have been enthusiastic about carbon pricing efforts: for the IMF, see Parry, Black, and Roaf (2021), Parry (2021), and Jessop, Eloraby, and Volcovici (2022); for the World Bank, see World Bank (2014); and for the OECD, see OECD (2021) and IMF and OECD (2021). Relatedly, Article 6 of the Paris Accord agreements of 2016 seeks to leverage trading to find low-cost approaches to mitigating emissions around the world. Article 6 governs the trading of carbon credits, allowing an entity in one country to pay for emissions reductions in another country. In this case, however, the emissions reductions may be undertaken voluntarily and may not reflect an explicit policy to price emissions (World Bank 2022; Edmonds et al. 2021).

Competitiveness Concerns with Carbon Pricing

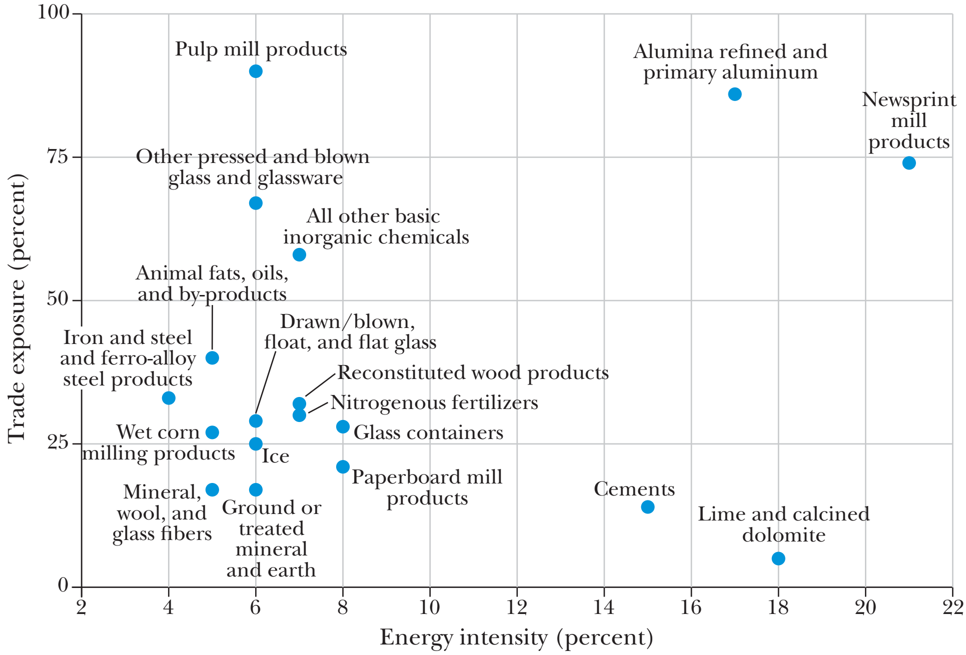
There are several domestic impediments to the adoption of emissions policies that impose costs. For example, one concern is the costs to households, particularly those lower in the income distribution. However, such concerns can be offset through other changes in the tax system, including by using carbon fee revenues to compensate lower-income taxpayers. These issues have been addressed elsewhere in the literature (for example, Goulder et al. 2019; Horowitz et al. 2017).

In this essay, we focus on domestic industry concerns that a carbon-pricing policy would injure their competitiveness. Consider a hypothetical jurisdiction that implements a carbon price of $110 per metric ton with no other policy response; its producers would compete with producers from other countries that may face no carbon price or that might even have their energy use subsidized. This would generate a competitive disadvantage in the hypothetical jurisdiction's local market, where imports may have cost advantages, and in markets abroad, where competitors may have lower costs of production.

The industries most exposed to competitiveness effects would be those with high energy-intensity and high exposure to trade. Figure 2 shows these industries, using data from the United States in 2019. Energy intensity on the horizontal axis is measured by industry fuel and electricity consumption scaled by industry level shipments. Trade exposure on the vertical axis is measured by total trade (exports plus

[2] Authors' calculations based on data from OECD (2022).

Figure 2
Energy Intensive and Trade Exposed Industries in the United States



Source: Authors' calculations based on US Census Bureau data for 2019.
Notes: Energy intensity on the horizontal axis is measured by industry fuel and electricity consumption scaled by industry level shipments. Trade exposure on the vertical axis is measured by total trade (exports plus imports) relative to total domestic shipments plus imports.

imports) relative to total domestic shipments plus imports. These data indicate that the industries most affected by such competitiveness concerns would be iron and steel, aluminum, newsprint, glass, and chemicals.[3]

To date, the most common policy approach to address competitiveness concerns has been to compensate industries for a large portion of their emissions, so that they only face the carbon price for marginal emissions. For example, in a cap and trade system, firms sometimes receive free permits that allow them to produce at prior production levels without facing an economic loss, but the ability to trade permits means that producers still face marginal incentives to reduce emissions that are analogous to a carbon price; every additional unit they emit costs them either the cost of a permit (if they need to buy one to reach their ideal production levels) or the opportunity cost of not selling permits at the going price (if they do not need to purchase permits at target production levels). These free permit allocations can

[3] We do not account for non-energy carbon emissions, which are present in meaningful amounts in the steel and cement industries.

then be reduced over time, through reductions in the total number of permits or other phase-outs. For example, California allocated free carbon permits to industry based on a formula that includes a facility's annual production and a benchmark emissions rate, as well as an adjustment factor that declines over time.

However, free grandfathering of allocations for carbon permits only imperfectly restores competitiveness to industries in jurisdictions that price carbon emissions. For instance, the free allocations cover an industry's direct carbon emissions but do not address the fact that energy inputs may be more expensive. Also, the number of free allocations may decline over time, imposing more costs on firms.

In these situations, multinational companies may find it advantageous to relocate carbon-intensive production to other countries. In addition, consumers may find it advantageous to purchase carbon-intensive imports offered at a lower price due to the absence of a carbon price. When behavioral responses like this reduce the amount of policy-induced global greenhouse gas emissions reduction, the effects are referred to as "carbon leakage."

The existing empirical estimates and model-based studies suggest that emissions leakage is limited in practice (Grubb et al. 2022), and a related literature has found inconsistent empirical evidence that firms move to "pollution havens" with low levels of environmental regulation (for example, Aldy and Pizer 2015; Singhania and Saini 2021; see also Levinson's article in this issue). Nonetheless, industries (such as those in Figure 2) that are both trade-intensive and energy-intensive are likely to face substantial concerns about this competitiveness channel.

In considering concerns about carbon pricing, it is important to recognize that more than 70 percent of greenhouse gas emissions are domestic, and issues of trade and competitiveness are far less important for addressing those sources of emissions. Of course, drawing a clean line between traded and nontraded sectors can be difficult. Goods that are not traded across national borders are often still influenced on the margin by conditions in international markets.

Subsidies for Investment and Innovation to Reduce Carbon Emissions

Some jurisdictions are reluctant to impose costs of carbon reduction directly on firms and consumers. This fear can lead to either inaction or to imposition of costs in other forms. As a prominent example, the United States (at the federal level) enacted a burst of spending on clean energy and innovation in 2021 and 2022, including spending for clean energy infrastructure and investments as well as a long list of clean energy tax credits.

The major piece of climate legislation was the 2022 Inflation Reduction Act. Estimates from the Joint Committee on Taxation and the Congressional Budget Office (2022) pegged the spending on clean energy tax credits and subsidies in legislation at more than $350 billion over ten years, although outside estimates suggest the fiscal costs could be substantially higher if the take-up of tax credits is higher than projected by the government (Bistline, Mehrotra, and Wolfram 2023;

Jiang et al. 2022; Penn-Wharton Budget Model 2023). The Infrastructure Investment and Jobs Act of 2021 contains additional clean energy infrastructure investments, including for electric vehicle charging and electricity transmission. Several studies have forecast the likely effectiveness of these subsidies in reducing carbon emissions. As summarized in Bistline et al. (2022), the studies find that the Inflation Reduction Act will likely reduce US carbon emissions by 32–42 percent below 2005 levels in 2030, an improvement relative to a baseline reduction of 6–11 percent without the legislation.

Figure 3 shows those countries with more than $1 billion in spending for clean energy over the period 2020–2021. While this gives some sense of the extent of recent subsidies, note that these figures are not scaled by country GDP, and some smaller economies spend more as a share of their economy than does the United States or other countries in this figure.

As with carbon pricing initiatives, subsidy policies vary both across and within jurisdictions. Some policies focus on nascent technologies; others subsidize the use of long-established technologies that might otherwise be phased out (like the tax credit in the Inflation Reduction Act of 2022 for existing nuclear production). Some policies spend directly on infrastructure like charging stations for electric vehicles, whereas other policies provide tax credits or grants for private actors based on their investment, production, or consumption. These policies all have one feature in common: they reduce the costs of investment and/or innovation for private market participants.

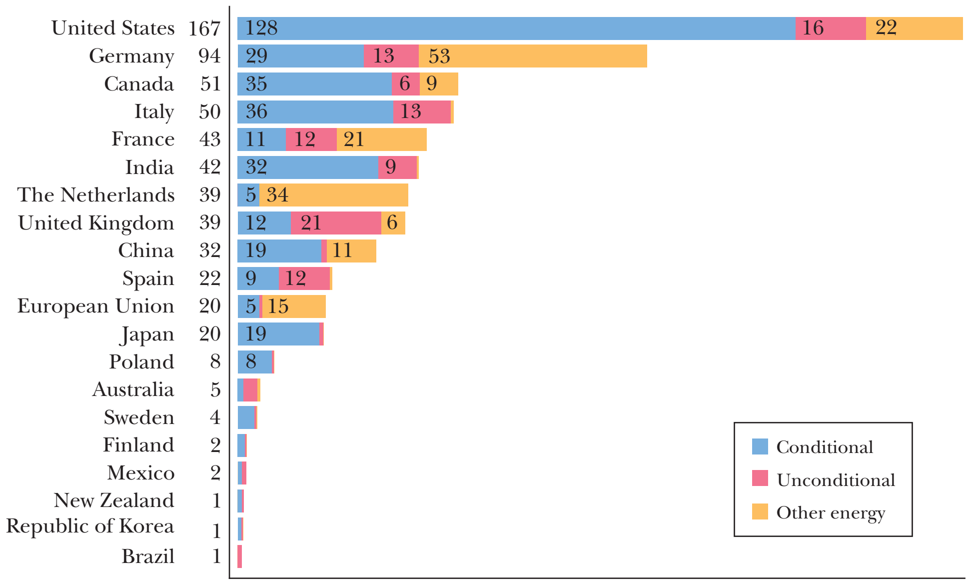
While subsidies are often chosen in part due to concerns about imposing costs on consumers or producers, it is important to remember that they have important distributional consequences of their own. For instance, the tax credits in the Inflation Reduction Act of 2022 will likely disproportionately benefit taxpayers that have higher incomes (Furman 2023). While subsidies avoid the prospect of directly harming those lower in the income distribution, they also have real fiscal consequences—including less revenue for alternative spending or tax cuts—that may be consequential for these taxpayers.

Finally, note that many jurisdictions pursue more than one type of policy. Some governments rely on both cost-imposing policies (such as taxes, fees, and regulation) as well as cost-reducing policies intended to spur clean energy production. An energy-intensive firm in such jurisdictions may find their fossil-fuel based energy costs rising even as renewable-sourced energy costs are falling. For example, many European countries subsidize the purchase of electric vehicles, support clean-tech manufacturing, and subsidize clean energy production (for more detail, see Kleimann et al. 2023).

Competitiveness Concerns and Subsidy Races

Subsidizing investments in carbon-free sources of energy can be economically efficient if free markets underprovide them. This is likely the case in the absence of a coordinated global response to the negative emissions externality, because emissions will still exceed their optimal level, and further reductions in

Figure 3
Public Investment in Clean Energy by Country and Policy Type, 2020–2021
(in billions of US dollars)



Source: Underlying data from group of nonprofit organizations that track public funding for energy. See: https://www.energypolicytracker.org/about/. The data only reflect subsidies implemented in 2020–2021; they do not capture ongoing programs, such as US subsidies under the Energy Policy Act of 2005, the precursor to the Inflation Reduction Act of 2002. They may count planned future spending in a single year. More generally, these data may not reflect longer term trends in spending on clean energy and tax subsidies. However, consistent cross-country data on these measures are not available for a longer time series, although there are initiatives to collect such data in the future.
Note: The blue "conditional" bars show policies that likely support the transition away from fossil fuels, but do not specifically require environmental safeguards, such as support for public transportation or electric vehicles. The red "unconditional" bars show policies that directly support production or consumption of low-carbon energy. The yellow "other energy" bars show other policies, including (but not limited to) those that support nuclear energy, biofuels, biomass, incineration, and hydrogen.

emissions will generate social benefits that exceed the private benefits. From this starting point, there can be enormous gains to the entire world from technological improvements that enable less expensive clean energy production, carbon capture and sequestration, and new technological innovations (Jaffe, Newell, and Stavins 2005). In addition, innovation in carbon mitigation technologies face the same market failures as any innovation: gains are unlikely to be fully captured by the private actors that undertake the relevant investments. This may be particularly true in nascent industries, where learning and technological advancement will lead to industry-wide cost reductions, generating external benefits that do not accrue to early producers. However, instead of strict protection of intellectual property, which is a typical policy lever to incentivize innovation in other contexts, it is important for governments to encourage knowledge transfer and diffusion of technologies that will help reduce carbon emissions. As one example, Athey et al. (2021) argue for mechanisms like advance market commitments for carbon removal technologies.

However, subsidies to reduce the costs of carbon-free energy have downsides. First, there are budgetary costs, which may be sustainable only in certain fiscal environments. Consider the United States, a useful case given the extent of the subsidies included in the Inflation Reduction Act of 2022, as well as the international reactions to them. The US Congress passed the Inflation Reduction Act of 2022 and the Infrastructure Investment and Jobs Act of 2021 in a relatively permissive fiscal environment, following years of low inflation and low interest rates. As interest burdens on the federal debt increase, alongside high ongoing deficits, fiscal constraints may become more binding.

Second, there are possible negative effects on other countries. For example, while subsidies to development of carbon-reducing energy sources in the United States can benefit other countries in important ways—including by improving technological progress in clean energy production and by reducing US greenhouse gas emissions—they also raise concerns. In the short run, US subsidies may attract investment, scarce expertise, and critical inputs for the energy transition away from other markets. In addition, US industries will be advantaged relative to those abroad if their energy costs are lower. While officials in the US government have urged other countries to also subsidize their energy transitions, many other countries may not be able to afford a subsidy-based approach, particularly low- and middle-income countries. Moreover, the marginal cost of public funds may be higher in many lower-income countries than in typical high-income countries, due to inefficiencies in tax collection (Besley and Persson 2014).

Further, foreign concerns about possible negative effects from US subsidies were magnified by the explicit inclusion of "domestic content" preferences in the US legislation; that is, multiple tax credits, including those for wind, solar, and electric vehicles, provided more favorable terms for products that were either made in the United States or, in the case of electric vehicles, in a country with whom the United States had a free trade agreement. While the latter inclusion may have mollified Canadian opposition to these provisions, other trading partners remained deeply concerned about losing production activities to subsidized locations. As one example, Tesla announced that it would move a battery manufacturing facility from Germany to the United States soon after the Inflation Reduction Act passed in 2022.

As a consequence of these concerns, the European Union, Japan, Korea, and the United Kingdom all complained vociferously about the domestic content provisions of the US clean energy subsidies (for news coverage, see Go 2022; Stangarone 2022; Parker, Bounds, and Williams 2022). In December 2022, France and Germany put forward a statement arguing that the US Inflation Reduction Act implies that Europe needs to adopt a more aggressive industrial policy (Le Maire and Habeck

2022). These issues were also raised at the highest levels, including during White House visits by French President Emmanuel Macron (in December 2022) and European Commission President Ursula von der Leyen (in March 2023). Countries have also begun to negotiate limited trade agreements to attain access to some of the credits, the first of which was between the United States and Japan in March 2023.

US policy stances have also raised concerns about broader effects on the international trading system. Global trade agreements have often focused on reducing domestic content rules and government subsidies to industry. These shifts in US policy worry governments that are in favor of a rules-based trading system, as US policy actions may foreshadow less restraint on these policy tools in broader arenas.

A third concern, related to the controversies over industrial policy, is that US subsidies for clean-energy industries could lead to a subsidy race. The terminology of a "race" refers to the idea of an arms race, when both sides would save resources by agreeing not to engage in such a race; it is one example of a prisoner's dilemma whereby global collective action can achieve better outcomes than individual jurisdictions operating noncooperatively. For example, in the short term, competition for scarce inputs or expertise may raise energy transition costs in other countries. In addition, subsidy races are expensive, putting substantial fiscal strains on countries that enter the race and often excluding lower income countries from the competition.

However, in this case, due to the positive externalities in clean energy sectors, it is not clear that a subsidy race in this specific area is always inefficient. Although there are elements of zero-sum competition, one country's subsidies for clean energy also have the potential to lower worldwide costs of clean energy adoption through industry-wide scale effects, and by leading to important technological innovation. As one example, Chinese subsidies to solar industry production served an important role in lowering the costs of solar energy, leading to greater solar adoption worldwide (Nemet 2019).

A final concern is that the emphasis on subsidies may affect support for price-based emissions policies, risking decreased support for carbon pricing or for removal of existing fossil fuel subsidies.[4] Countries that impose carbon prices may feel the need to join a subsidy race rather than imposing costs on their producers. In addition, the passage of the US subsidy-based climate policy has been taken by some as an argument that carbon pricing is not necessary or desirable in the US context (Kaufman 2023).

[4] The size of US fossil fuel subsidies is significant. Tax measures alone generate a fiscal cost of more than $31 billion over ten years (US Department of the Treasury 2023, p. 213).

Carbon Border Adjustment Mechanisms

A "carbon border adjustment mechanism" describes a policy where a jurisdiction with carbon pricing applies import fees based on the carbon content of imported goods—the amount of the border adjustment fee is based on the local carbon price, with an adjustment for any carbon price in the exporting country. For example, if the home market has a $110 per ton carbon price and the foreign market has a $10 per ton carbon price, the tariff would be $100 per ton of carbon embedded in the product. If one unit of the product contained the equivalent of 0.02 tons of carbon, that would imply a $2 fee per unit.

Many competitiveness issues raised by a carbon-pricing policy in the domestic market are addressed by the carbon border adjustment mechanism. In countries with a carbon border adjustment mechanism, all consumption goods face the same costs associated with their carbon emissions.[5]

In December 2022, the European Parliament and the Council of the European Union reached a provisional agreement to implement a carbon border adjustment mechanism beginning in late 2023; this agreement was finalized in April 2023. The proposed mechanism would levy a fee on imported goods in key energy-intensive, trade-exposed industries; the fee would be set at a level that would offset the competitive disadvantage associated with European Union carbon costs.[6] The proposal was designed to level the playing field in a manner that was consistent with World Trade Organization rules. For example, the EU carbon border adjustment mechanism would be phased in as free allowances for carbon-based output are phased out, causing both domestic and foreign producers to be treated similarly. Canada and the United Kingdom are also considering implementing carbon border adjustment mechanisms.

While a carbon border adjustment can address competitiveness issues in the local market, there are also important questions about competitiveness in external markets. One approach would be to refund the domestic carbon prices for exports. While this kind of export rebate can address competitiveness in third markets, it is likely to prove contentious. Rebating carbon fees for exports runs the risk of dampening emissions reduction efforts at home. It may also raise political concerns to treat carbon produced for export goods differently from carbon involved in domestic production.

[5] The argument for a carbon border adjustment mechanism is different from arguments that would equalize other policy or economic differences among countries, by (for example) levying tariffs when foreign minimum wages (or foreign wages in general) are lower. Countries set their own minimum wage laws in ways that are sensitive to circumstances, and wages in poorer countries are lower for a variety of factors that ultimately reflect lower economic productivity in poorer countries. Most important, unlike greenhouse gas emissions, these labor market differences do not typically generate global market externalities, so there are fewer concerns regarding international policy spillovers, such as leakage and free-riding.

[6] An analysis of the proposed carbon border adjustment mechanism in 2021 showed that four industries—iron and steel, cement and lime, fertilizer, and aluminum—account for about 55 percent of European Union industrial carbon-equivalent emissions, which themselves are about 25 percent of European Union carbon equivalent emissions (European Commission 2021, Figure 7). Note that the proposed mechanism has been expanded beyond the sectors analyzed in the Commission report.

Measurement and Mitigation Incentives

Implementing a carbon border adjustment mechanism poses practical challenges. One central question is measurement. A carbon border adjustment mechanism is based on the carbon-content of individual imported goods, adjusting for the cost difference between domestic and foreign carbon costs.

As one might expect, informational imperfections make it difficult to assess the carbon-intensity of individual shipments. If customs officials instead rely on more aggregated measures, like industry- or country-level emissions ratios, it can affect incentives in unexpected or even counterproductive ways. For example, imagine the European Union levied a fee based on the carbon-intensity of foreign aluminum imports measured at the plant level. Exporters facing levies would have an incentive to send aluminum from their cleanest (least carbon intensive) plants to Europe, while sending the dirtier products to other countries, a process described as "reshuffling" (Fowlie, Petersen, and Reguant 2021).

If the measurement were instead done at the industry level, that would reduce the incentive to reshuffle, but it would also dampen the incentives of individual producers to reduce their carbon emissions—because they would simply be assigned the industry-average assessment regardless. One possible way around the latter problem would be to allow companies to opt for self-certification, charging the remaining companies based on the average emissions of the residual group of companies. Assuming one can measure and monitor both company and total industry emissions, this approach has the potential to lead to efficiency gains, as described in Cicala, Hémous, and Olsen (2022).

More generally, one can imagine a carbon border adjustment mechanism leading to broad reshuffling, whereby dirty exports head to countries without carbon border adjustment mechanism at a lower price (potentially increasing demand for such products in those markets) and clean country exports serve the markets of countries with a carbon border adjustment mechanism. In these instances, the aggregate impact on emissions is likely to be small.

If measurement of carbon content is done at a more aggregate level, policymakers may take mitigation actions in order to reduce the tariffs faced by a subset of industries; we discuss these policy dynamics in more detail below. Still, although carbon border adjustment mechanisms address competitiveness concerns, their effect on emissions will quantitatively depend on how important energy-intensive export markets are for the trading partners of carbon border adjustment countries. The literature (drawing mainly on simulations) tends to find that they have only small effects on emissions, production, and welfare (Böhringer, Balistreri, and Rutherford 2012; Devarajan et al. 2022; Irfanoglu, Sesmero, and Golub 2015; Branger and Quirion 2014).

Supply Chain Complications

A carbon border mechanism also poses vexing issues surrounding value chains. While carbon border adjustment mechanisms have the potential to address competitiveness issues faced by industries in cost-imposing jurisdictions, they do not address the competitiveness of industries that use those products intensively as inputs.

For example, if the European Union adopts its proposed carbon border adjustment mechanism, it will adjust for differences in the carbon price applied to steel imports, but not for automobile imports, at least for the first several years. To the extent EU automobile manufacturers face higher steel prices, they would be at a competitive disadvantage unaddressed by the carbon border adjustment mechanism. Given that many of the most-carbon intensive products (iron/steel, aluminum, chemicals, glass, fertilizers, and so on) can be important inputs for downstream industries, it may prove difficult to completely level the competitiveness playing field in the absence of an economy-wide carbon price and border adjustment.

Carbon Border Adjustments without Carbon Pricing?

Countries that rely on subsidies as the central element of their climate policy do not face the competitiveness concerns that arise from domestic carbon pricing. Their industries do not face increased costs due to government climate change mitigation policies; on the contrary, energy costs paid by industry may be reduced by government subsidies. For example, the United States does not currently impose a nationwide carbon price, and while some US states impose costs on their firms, these jurisdictions do not generally host many firms in energy-intensive, traded industries. For example, examining US sources of production of the products targeted for inclusion in the proposed EU carbon border adjustment mechanism—like steel, aluminum, cement and chemicals—the implied average US carbon price is very low, under $1 per ton.

However, once carbon border adjustments are part of the political discussion, domestic actors could use them as an excuse to seek protectionism, even in cases where little "adjustment" is actually required. For example, US industries could claim that their low emissions alone should justify a border adjustment, even though their industries do not face carbon costs and may even benefit from energy subsidies that lower their input costs. The United States government has made a proposal along these lines for the steel industry, working toward a "green steel deal" with the European Union (as reported in Swanson 2022).[7] However, a carbon border adjustment mechanism in such instances might usefully be relabelled as a "carbon tariff."

Such carbon tariffs are likely to be perceived as unfair abroad, by two sets of jurisdictions. For those jurisdictions that are imposing carbon costs and imposing

[7] The US proposal envisions both the United States and the European Union (alongside other countries that might join the group) levying tariffs on emissions-intensive steel from other countries outside the group (Swanson 2022). However, the steel arrangement has somewhat unique preconditions. For instance, the United States has long protected the steel industry with tariffs through antidumping provisions and, in 2017, the unfounded invocation of Section 232, which allowed a levying of broad tariffs on steel and aluminum on national security grounds. In December 2022, the World Trade Organization ruled that these tariffs were not consistent with trade rules, but the US Trade Representative stated that the World Trade Organization should not judge US national security interests.

a carbon border adjustment mechanism as a consequence, a carbon tariff would keep intact any competitive disadvantage faced by their producers relative to US producers. This could unravel efficient policies, by lowering support for cost-imposition abroad, as foreign producers will suspect (justifiably) that a level playing field is impossible. Such producers may seek countervailing subsidies under the argument that subsidies are required to maintain fair competition. Such a dynamic could unravel efficient climate policies and result in timid actions by governments that have little fiscal space for bold programs of subsidization. Other jurisdictions that do not impose their own carbon-pricing regimes may also complain about a US carbon tariff, arguing that such tariffs are little more than protectionism in disguise, given that the US government imposes no carbon price of its own. Finally, a US carbon tariff would do nothing to incentivize further reductions in US emissions.

For these reasons, economic logic suggests only using a carbon border adjustment mechanism in the presence of cost-imposing policies; such limits would also enable the carbon border adjustment mechanism to be implemented in a manner that was consistent with world trade obligations. While consistency with these rules has admittedly not been central to the policy-making concerns of US policymakers since 2017, the principles behind World Trade Organization rules nonetheless remain an important source of stability and fairness in the international trading system.

Domestic Policy Dynamics of Carbon Border Adjustments

Perhaps the most intriguing feature of carbon border adjustment mechanisms is not how they shape company decisions about emissions or even the patterns of international trade, but rather how they shape the policy choices of governments that determine the future path of greenhouse gas emissions. The ability to create positive policy spillovers may be a first-order determinant of ambitious climate policy.

In the United States, a border adjustment could facilitate the adoption of cost-imposing policies by addressing concerns regarding the erosion of domestic industry competitiveness, carbon leakage, and the free-riding of other countries. Indeed, carbon border adjustment has the potential to harness protectionist sentiment toward efficient ends. For example, the steel industry has frequently been successful in seeking tariff protection in recent decades. Some of the rationale for these tariffs has not been well-founded, including recent reliance on national security rationale. But in the presence of carbon pricing, a carbon border adjustment could serve efficiency by equalizing the costs associated with carbon emissions for all producers serving the US market.

In fact, many US industries could gain competitiveness from a price-based approach that included a border adjustment, because US industrial production tends to be less carbon-intensive than that of several of our largest trading partners (Rorke and Bertelsen 2020). One report suggested that a $43 per ton carbon tax and accompanying border adjustment may cause US imports to fall considerably, while U.S. steel industry output would expand (CRU Consulting 2021).

A carbon border adjustment mechanism might also enable a useful pivot toward pricing tools as a complement to subsidies in the US government approach to climate mitigation. While most models predict that the US subsidy-based approach to climate policy will be somewhat effective, they also predict that this approach will not lead to sufficient emissions reductions, particularly in the industrial sector, strengthening the rationale for at least some carbon price on industrial emissions (Bistline, Mehrotra, and Wolfram 2023). Timilsina (2022) reviews numerous studies that speak to the greater efficiency of price-based mitigation policies.

Moreover fiscal constraints may make it attractive to use carbon pricing as a complementary approach to subsidies. The two policies together can achieve greater emissions reductions at a lower fiscal cost, while also providing revenue to insulate households from increased costs (Roy, Burtraw, and Rennert 2021).

Finally, the clean energy subsidies may change the political economy of price-based approaches, by increasing the size and power of industrial sectors that would also benefit from carbon pricing (including wind, solar, electric vehicles, batteries, nuclear power, carbon sequestration, sustainable aviation fuel, and others), while shrinking the power and market size of the fossil fuel industry. In the end, cost-increasing and cost-reducing policies may be complementary policy instruments. For instance, the European Union has long subsidized the development of clean electricity production, and this groundwork helped enable a stronger carbon pricing system.

Policy Dynamics of Carbon Border Adjustments Abroad

More generally, if carbon border adjustment mechanisms were broadly applied by a wide group of importers, such mechanisms have the potential to induce virtuous policy changes abroad for several reasons. First, if the jurisdiction in question is dependent on carbon-intensive exports to countries that are imposing a carbon border adjustment mechanism, it may find adoption of symmetric carbon pricing (which would eliminate the tariff) advantageous, or it may increase an existing carbon price to lessen the tariff (Böhringer, Carbone, and Rutherford 2016). Even sector-level carbon pricing may be sufficient to turn off the tariff. Second, increased carbon pricing would have the benefit of converting foreign tariff revenue to domestic revenue. If a domestic company has to pay for its carbon content when shipping to a country with a carbon border adjustment mechanism, its payments might as well instead benefit the domestic treasury. Both of these effects quantitatively depend on how important energy-intensive export markets are for the trading partners of countries that are imposing a carbon border adjustment mechanism.

Third, in exporting countries, a foreign carbon border adjustment mechanism could provide political cover and rhetorical arguments for making the transition to cost-imposing policies. For example, Türkiye, which sends nearly half of its exports to the European Union, has considered imposing its own carbon price in response to the EU carbon border adjustment mechanism (Weise 2021). Similarly, the EU adjustment mechanism has also been credited with pushing Russia to announce a carbon neutrality goal and experiment with carbon pricing—before Russia's invasion of Ukraine scrambled trade and political relations (Zabanova 2021). The European Union adjusted the regulations for its proposed carbon border adjustment mechanism to address concerns that countries would implement carbon prices that were only assessed on exports to the European Union, deeming such a scheme a circumvention and ineligible for credit towards the carbon border adjustment mechanism.

Finally, if there is widespread adoption of a carbon border adjustment mechanism, there may also be a symbolic or moral rationale for implementing carbon pricing to qualify for the "in" group and avoid barriers. While the strength of this motive should not be overstated, it might influence those countries that want to be seen as good actors with respect to climate policy.

However, the foreign policy responses to a carbon border adjustment mechanism need not be accommodating. Not all governments abroad will be sanguine about the threat of foreign tariffs; for example, China, India, Indonesia, and Thailand have suggested that they will oppose carbon border adjustment mechanisms on the grounds that they are protectionist and discriminatory (Böhringer et al. 2022). If some countries launch retaliatory actions in response, this risks a trade war of escalating tariffs, reducing the gains from trade on both sides and harming international cooperative efforts on both climate and other areas of joint concern. Given the many agenda items requiring international cooperation (including security, public health, and tax competition), the downsides of additional trade frictions are substantial, especially at a time when the international trading system is already under strain.

While we have highlighted how the economic argument for carbon border adjustment mechanisms is strongest in the presence of cost-imposing policies, this choice is not strictly dichotomous. Policymakers could choose to adjust the share of domestic emissions subject to a carbon price. For example, a bill proposed in the US Senate would charge a carbon price on US plants whose emissions are above some threshold (currently set at US average emissions), and a corresponding carbon border adjustment mechanism on imports for their carbon emissions above that threshold.[8] This approach does not resolve the competitiveness concerns of other jurisdictions that impose costs on all carbon emissions, but it does have the advantage of treating all emissions similarly in the domestic market. In addition, the threshold itself is an important policy dial. As the threshold increases, this approach mimics a carbon tariff with no corresponding domestic cost, since costs are mostly imposed on carbon-intensive imports rather than domestic production. As the threshold goes down and approaches zero, the policy approaches an

[8] The Clean Competition Act (S.4355 from the 2021–2022 U.S. Congress) is described at https://www.congress.gov/bill/117th-congress/senate-bill/4355. CRU Consulting (2021) reports that steel production in China, the world's largest exporter, are 1.8 and 5 times more emissions intensive than US steel production for flat and long steel products, respectively; thus, carbon fees would be higher for Chinese producers than those in the United States.

industry-specific carbon price applying to domestic producers and with an accompanying carbon border adjustment mechanism.

Climate Clubs

Some countries undertake ambitious and costly policies in order to reduce greenhouse gas emissions; others do not. Since climate change is a global phenomenon, no jurisdiction will internalize the externalities associated with greenhouse gas emissions, given that the vast majority of the benefits from emissions reduction efforts benefit those outside their borders. Absent a coordinating mechanism, jurisdictions have a self-interested incentive to do suboptimal amounts of mitigation, leading to a free-riding effect.

For almost 30 years, the main international coordinating mechanism has been the periodic meetings of the UN Framework Convention on Climate Change. However, these meetings have emphasized voluntary pledges for reducing carbon emissions, including "nationally determined contributions," a concept that was central to the Paris Accord agreements adopted in 2016. There is no enforcement mechanism to address countries that either commit too little or do not meet their commitments. Since the Paris Accord, countries have fallen woefully short, both making commitments that are not sufficiently ambitious and failing to enact policies that meet existing commitments (Climate Action Tracker 2022).

This free-rider concern led Nordhaus (2015) to propose a climate club, whereby ambitious jurisdictions could group together and seek to prevent free-riding behavior by levying penalties on insufficiently ambitious jurisdictions. In theory, ambition need not take the form of a carbon price: other measures that reduce emissions could be considered equivalent. In practice, Nordhaus cautioned that relying on price mechanisms would make implementation far more straightforward. Otherwise, it would be difficult to measure the relative stringency of countries' policies, and any such judgements could swiftly become contentious.

In a climate club, penalties against low-ambition countries could take the form of a broad tariff on all imports, which has certain advantages relative to a carbon border adjustment mechanism. It is administratively simpler, because there is no need to measure the carbon content of imports. It is also less prone to trade reshuffling in response, because all goods from nonclub countries would face the tariff.

Differences from a Carbon Border Adjustment Mechanism

Climate clubs differ from carbon border adjustments in that they are not meant to level the playing field for any particular good. Instead, they seek to encourage policy ambition by penalizing insufficiently ambitious countries with an across-the-board tariff. Like carbon border adjustments, they are capable of channeling protectionist sentiment toward potentially helpful ends, but there are also risks that the climate would be used as an excuse to impose tariffs that are not justified on climate policy grounds.

Climate clubs don't have generalizable fiscal implications. If countries adopt carbon pricing to join the club, that will generate substantial revenues for those countries, but choosing to pursue an ambitious program of subsidies to join the club would have the opposite budgetary effect. Carbon border adjustments are designed to fall on only a small number of sectors that are carbon-intensive in production, so they have relatively minor revenue impacts. However, a climate club could conceivably enact broader tariffs, and thus would generally raise more revenue than a carbon border adjustment mechanism—though far less than an economy-wide carbon price.

By the same logic, climate clubs should have a stronger incentive effect on "out" countries to enact climate policies, because nonadoption comes with larger trade barriers that affect a country's entire export sector, not just carbon-intensive products. Also, the rhetorical or symbolic motives for joining the climate club could also be stronger, as the club would be more explicitly designed to distinguish "good" (climate ambitious) countries from "bad" (free-riding) countries.

Challenges of Climate Clubs

Climate clubs face a number of practical challenges and concerns. First, it is likely to be difficult for countries to agree on how to measure the rough equivalence of policy ambition. Some countries may prefer a carbon tax while others prefer a cap-and-trade approach, and countries' carbon price choices are likely to differ. Further, some national economies may already be emitting less carbon per person, perhaps as a result of their industrial mix or because they already have a substantial amount of hydroelectric or nuclear power. Such countries may feel that their current policies are already sufficiently ambitious, despite limited new policy action. Further, in a real-world situation, membership "in" or "out" of a climate club will likely reflect political power and alliances, not just dispassionate measures of real climate policy action.

Perhaps the most difficult problems arise if some countries wish to certify their membership in the climate club by subsidies for green energy rather than carbon pricing. For example, the United States might seek to claim its place as a "high ambition country" based on the vast expenditures on clean energy tax credits and investments in the 2021 and 2022 legislation. But in this situation, the climate club does nothing to address policy concerns about competitiveness. For example, within the set of high-ambition countries, cost-reducing locations will have an advantage. Because climate clubs would implement a broad-based tariff without considering underlying policy differences across "in" countries, they cannot remedy industry competitiveness concerns (for trade-exposed, energy-intensive industries) absent much greater policy harmonization.

Indeed, in the presence of heterogeneous policy choices, where some jurisdictions impose costs and others subsidize, it is not possible to address both types of policy spillovers (competitiveness and free-riding) with a single remedy. A carbon border adjustment will not completely address free-riding, and a climate club will not address competitiveness.

Second, because the goal of a climate club is not to equalize policy-induced costs in particular industries (like a carbon border adjustment mechanism), the appropriate tariff level to impose outside the club becomes a political judgment. It does not take much imagination to realize that such judgements could swiftly become fraught. On a related note, if tariffs are significant, a climate club made up of primarily higher-income countries imposing broad tariffs on lower-income countries would strike many as punitive or unfair, which could weaken the moral impetus to join. These issues are discussed in the following section.

Third, countries outside of the climate club may respond by launching retaliatory trade measures, which could spark counter-retaliation and a trade war. These incentives would be even larger with a climate club than with a carbon border adjustment mechanism, due to the larger impact of the tariffs. In the extreme, a climate club could bifurcate the world such that countries in the club primarily trade with one another and countries outside the club likewise primarily trade with one another.

Finally, the rules of a climate club may also be hard to reconcile with commitments under the World Trade Organization agreements. Unlike a carbon border adjustment, the tariffs in a climate club are not designed to treat both foreign and domestic producers alike in the home market. While some argue that climate issues should be put before the arcane details of trade rules (for example, Rodrik 2022), it is important to avoid trade tensions that might ultimately risk both environmental objectives as well as the gains from trade. The challenge would be to modernize World Trade Organization rules to allow countries the freedom to take heterogenous emissions reduction strategies, without undermining the long-held objectives of the world trading system.

In theory, both carbon border adjustment mechanisms and climate clubs could operate without any tariffs actually coming into effect. Indeed, the ideal outcome would be for policy adoption abroad to forestall the use of tariffs, leading to an upward harmonization of climate policy. In that event, the threat of tariffs would remain just that. The example of Türkiye (described above) provides one real-world illustration of this process at work. But it remains unclear if such a “leveling-up” scenario is realistic.

Implications for Low-Income Countries

Poorer countries tend to face greater risks from climate change, as simulations show emerging markets (on average) bear higher relative costs in terms of economic disruption and loss of life (for example, Carleton et al. 2022). Poorer countries also frequently lack the resources and fiscal space to undertake climate change mitigation efforts, and given the opportunity costs of fiscal resources in low-income countries, their efforts often entail greater absolute levels of sacrifice. Finally, the existing stock of greenhouse gases in the atmosphere was mostly caused by economic activities in richer countries, so there is an important fairness argument that higher-income countries should bear the brunt of the costs.

Both carbon border adjustment mechanisms and climate clubs risk harming poorer countries in the event that their goods face tariffs abroad. Opportunities for export-led growth will be diminished, and tariffs will reduce the gains from trade. Böhringer, Carbone, and Rutherford (2018) show that carbon border adjustment mechanisms have important distributional consequences, reallocating abatement costs toward those countries facing levies. These considerations raise questions about the price of admission to the climate club, or what policy actions might be required in order to turn off a carbon border adjustment mechanism or club tariff.

One option is to require less policy action from poorer countries. IMF staff suggest a carbon pricing floor that varies with level of development (Parry, Black, and Roaf 2021). One scenario they consider would require price floors of $75 per ton for advanced economies, $50 per ton for higher-income emerging economies, and $25 per ton for low-income emerging economies—they show how these policies could complement existing "nationally determined contribution" commitments under the Paris agreement.

One could also imagine exempting the poorest countries from carbon border adjustment mechanisms or club tariffs, alongside earmarking revenues from border measures for a fund targeting emissions reductions in poorer countries. Since low-income country emissions are a relatively small fraction of the world total, significant gains can be achieved even when exempting the poorest countries. The 28 countries that the World Bank classifies as low-income generate 4 percent of world carbon emissions; even the 82 countries that are either low-income or lower-middle income (Venezuela was not reported) generate only about 25 percent of world emissions.

Poor countries have a lot to gain from successful international cooperation around emissions mitigation. In addition to bearing significant costs from inaction, they stand to benefit from the cost reductions and technological innovation associated with the clean energy transition, enabling their own emissions reductions to be done at lower cost when the time comes. Cost-imposing measures abroad are no threat to their competitiveness. While cost-reducing measures could give some foreign industries an advantage, they also generate greater scale and innovation in clean energy sectors.

Discussion

Climate policies are unsurprisingly heterogeneous. National economies specialize in different industries and generate wide-ranging standards of living, and their governments face varying political constraints, fiscal constraints, and circumstances. When countries' climate policies vary, those policies generate spillovers. Cost-imposing jurisdictions fear carbon leakage and negative competitiveness effects, and high-ambition countries fear that low-ambition countries will free-ride on their sacrifices.

International trade plays an important role in national decision-making about climate change mitigation efforts. About 25 percent of all greenhouse gas emissions are embodied in traded goods, and an economy's carbon "footprint" may differ substantially based on whether it is measured in production terms (as is typical) or in consumption terms (Wiedmann and Lenzen 2018). For example, China's economy produces more carbon-intensive products than it consumes, whereas the opposite is true for the United States. On average, lower-income countries produce more carbon-intensive products than they consume, while higher-income countries consume more carbon-intensive products than they produce (Liu et al. 2020; Zhu et al. 2018; Wood et al. 2020). International trade also has more general effects on carbon emissions, by altering the scale and composition of production as well as the spread of innovation (Copeland and Taylor 2004; Copeland, Shapiro, and Taylor 2022).

Ideally, countries would coordinate border measures in a way that encourages positive policy action rather than conflict. An ideal border adjustment or climate club would end up levying few tariffs, instead urging trading partners to respond to enforcement mechanisms with greater mitigation efforts. At present, four jurisdictions—the United States, China, India, and the European Union—account for about one-half of world carbon emissions. Thus, any coordination effort should pay close attention to the incentives of these jurisdictions, while also building a system that can address wide-ranging economic circumstances across the globe.

Trade negotiations can potentially serve as invaluable tools to further climate aims. The current trade policy structure works against climate change mitigation, since trade barriers (both tariff and nontariff) are far higher for low-carbon industries than for high-carbon industries. This provides large implicit subsidies to emissions-intensive production, relative to cleaner production (Shapiro 2021). New rounds of trade negotiation can aim to correct these perverse incentives, while lowering or eliminating trade barriers on goods, services, and technology that are needed to support clean energy adoption and innovation. Even if carbon border adjustment mechanisms or climate clubs result in some tariff increases, this can be done alongside broader efforts at green trade liberalization and a rebalancing of current tariffs structures.

International cooperation on climate mitigation policy can be enhanced through carrots as well as sticks (Jakob et al. 2022). Market access, including access to scarce supplies of key raw materials needed to produce clean energy, is an important carrot that can work alongside carbon border adjustments (or other sticks) to encourage countries to participate in cooperative solutions. In addition, subsidizing countries should commit to ensuring that the benefits of their investments in clean energy help the world adopt cleaner technologies, by working to facilitate knowledge spillovers across national boundaries through reduced barriers on clean energy trade and investment. This vision offers opportunities to make progress on climate without undermining the world trading system, which has generated enormous gains benefitting billions of people. Beyond these reforms, the World Trade Organization can continue to serve its enduring purpose: allowing countries to access the myriad gains from trade by ensuring predictable rules and collaborative solutions to global collective action problems.

These central themes—the inevitability of diverse policy actions, the nature of policy spillovers, and the ideal policy responses to address such spillovers—will define the world's ability to address climate change in an orderly fashion.

■ *We received helpful feedback from Joseph Aldy, Laura Alfaro, Chad Bown, Meredith Fowlie, Steven Fries, Luis Garicano, Cullen Hendrix, Arik Levinson, Thornton Matheson, Jean Pisani-Ferry, William Pizer, Jacquelyn Pless, André Sapir, Joseph Shapiro, Robert Stavins, Matthew Tarduno, Jeromin Zettelmeyer, and participants at the Harvard Kennedy School of Government's conference on the Political Economy of Climate Change Policy, the Peterson Institute for International Economics seminar, the Salata Institute for Climate and Sustainability at Harvard University conference on "Can Trade Policy Work for the Climate?", the National Tax Association Spring Symposium, and the World Resources Institute seminar. We received helpful research assistance from Andrew Chu, Julieta Contreras, James Murray, and Matthew Tarduno. We are grateful for helpful editorial suggestions and feedback from Erik Hurst, Nina Pavcnik, Timothy Taylor, and Heidi Williams. The PIIE research assistance for this project was funded by financial support from a multiyear grant from the Danish Embassy in Washington for PIIE research related to developing economies and climate change.*

References

Aldy, Joseph E., and William A. Pizer. 2015. "The Competitiveness Impacts of Climate Change Mitigation Policies." *Journal of the Association of Environmental and Resource Economists* 2 (4): 565–95.

Athey, Susan, Rachel Glennerster, Nan Ransohoff, and Christopher Snyder. 2021. "Advance Market Commitments Worked for Vaccines. They Could Work for Carbon Removal, too." *Politico*, December 21. https://www.politico.com/news/agenda/2021/12/22/carbon-removal-advance-market-commitments-525988.

Besley, Timothy, and Torsten Persson. 2014. "Why Do Developing Countries Tax So Little?" *Journal of Economic Perspectives* 28 (4): 99–120.

Bistline, John, Geoffrey Blanford, Maxwell Brown, Dallas Burtraw, Maya Domeshek, Jamil Farbes, Allen Fawcett et al. 2022. "Emissions and Energy System Impacts of the Inflation Reduction Act of 2022." Unpublished.

Bistline, John, Neil R. Mehrotra, and Catherine Wolfram. 2023. "Economic Implications of the Climate Provisions in the Inflation Reduction Act." *Brookings Papers on Economic Activity*. https://www.brookings.edu/bpea-articles/economic-implications-of-the-climate-provisions-of-the-inflation-reduction-act/.

Böhringer, Christoph, Edward J. Balistreri, and Thomas F. Rutherford. 2012. "The Role of Border Carbon Adjustment in Unilateral Climate Policy: Overview of an Energy Modeling Forum Study (EMF 29)." *Energy Economics* 34 (S2): S97–110.

Böhringer, Christoph, Jared C. Carbone, and Thomas F. Rutherford. 2016. "The Strategic Value of Carbon Tariffs." *American Economic Journal: Economic Policy* 8 (1): 28–51.

Böhringer, Christoph, Jared C. Carbone, and Thomas F. Rutherford. 2018. "Embodied Carbon Tariffs." *Scandinavian Journal of Economics* 120 (1): 183–210.

Böhringer, Christoph, Carolyn Fischer, Knut Einar Rosendahl, and Thomas Fox Rutherford. 2022. "Potential Impacts and Challenges of Border Carbon Adjustments." *Nature Climate Change* 12 (1): 22–29.

Branger, Frédéric, and Philippe Quirion. 2014. "Would Border Carbon Adjustments Prevent Carbon Leakage and Heavy Industry Competitiveness Losses? Insights from a Meta-analysis of Recent Economic Studies." *Ecological Economics* 99: 29–39.

Carleton, Tamma, Amir Jina, Michael Delgado, Michael Greenstone, Trevor Houser, Solomon Hsiang, Andrew Hultgren et al. 2022. "Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits." *Quarterly Journal of Economics* 137 (4): 2037–2105.

Cicala, Steven, David Hémous, and Morton Olsen. 2022. "Adverse Selection as a Policy Instrument: Unraveling Climate Change." NBER Working Paper 30283.

Clausing, Kimberly A., and Catherine Wolfram. 2023. "Replication data for: Carbon Border Adjustments, Climate Clubs, and Subsidy Races When Climate Policies Vary." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E192048V1.

Climate Action Tracker. 2022. "CAT Emissions Gap." Climate Analytics and NewClimate Institute. https://climateactiontracker.org/global/cat-emissions-gaps/.

Congressional Budget Office. 2022. "Congressional Budget Office Cost Estimate." https://www.cbo.gov/system/files/2022-08/hr5376_IR_Act_8-3-22.pdf.

Copeland, Brian R., Joseph S. Shapiro, and M. Scott Taylor. 2022. "Globalization and the Environment." In *Handbook of International Economics: International Trade*, Vol. 5, edited by Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff, 61–146.

Copeland, Brian R., and M. Scott Taylor. 2004. "Trade, Growth, and the Environment." *Journal of Economic Literature* 42 (1): 7–71.

CRU Consulting. 2021. *Leveraging a Carbon Advantage: Impacts of a Border Carbon Adjustment and Carbon Fee on the US Steel Industry*. London: CRU International Ltd.

Devarajan, Shantayanan, Delfin S. Go, Sherman Robinson, and Karen Thierfelder. 2022. "How Carbon Tariffs and Climate Change Can Slow Global Warming." PIIE Working Paper 22-14.

Edmonds, James, Sha Yu, Haewon Mcjeon, Dirk Forrister, Joseph Aldy, Nathan Hultman, Ryna Cui et al. 2021. "How Much Could Article 6 Enhance Nationally Determined Contribution Ambition toward Paris Agreement Goals through Economic Efficiency?" *Climate Change Economics* 12 (2): 2150007.

Energy Policy Tracker. 2020–2021. "Country Policies." https://www.energypolicytracker.org/policies-analysis# (accessed May 2023).

European Commission. 2021. *Proposal for a Regulation of the European Parliament and of the Council: Establishing a Carbon Border Adjustment Mechanism*. 2021/0214 (COD). Brussels: European Commission.

European Environment Agency. 1990–2021. "EEA greenhouse gases—data viewer." https://www.eea.europa.eu/data-and-maps/data/data-viewers/greenhouse-gases-viewer (accessed May 2023).

Fowlie, Meredith, Claire Petersen, and Mar Reguant. 2021. "Border Carbon Adjustments When Carbon Intensity Varies across Producers: Evidence from California." *AEA Papers and Proceedings* 111: 401–05.

Furman, Jason. 2023. "Comment on 'Economic Implications of the Climate Provisions in the Inflation Reduction Act'." *Brookings Papers on Economic Activity*. https://www.brookings.edu/wp-content/uploads/2023/02/2b_20230330-BPEA-climate-furman-comment.pdf.

Go, Jiyeong. 2022. "EU and Japan Lash Out at 'Discriminatory' US Green Subsidies." f*Di Intelligence* (*Financial Times*), November 15. https://www.fdiintelligence.com/content/news/eu-and-japan-lash-out-at-discriminatoryus-green-subsidies-81649.

Goulder, Lawrence H., Marc A. C. Hafstead, GyuRim Kim, and Xianling Long. 2019. "Impacts of a Carbon Tax across US Household Income Groups: What Are the Equity-Efficiency Trade-Offs?" *Journal of Public Economics* 175: 44–64.

Grubb, Michael, Nino David Jordan, Edgar Hertwich, Karsten Neuhoff, Kasturi Das, Kaushik Ranjan Bandyopadhyay, Harro van Asselt et al. 2022. "Carbon Leakage, Consumption, and Trade." *Annual Review of Environment and Resources* 47: 753–95.

Horowitz, John, Julie-Anne Cronin, Hannah Hawkins, Laura Konda, and Alex Yuskavage. 2017. "Methodology for Analyzing a Carbon Tax." Office of Tax Analysis Working Paper 115.

IMF, and OECD. 2021. *Tax Policy and Climate Change: IMF/OECD Report for the G20 Finance Ministers and Central Bank Governors*. Italy: IMF and OECD.

Irfanoglu, Zeynep Burcu, Juan P. Sesmero, and Alla Golub. 2015. "Potential of Border Tax Adjustments to Deter Free Riding in International Climate Agreements." *Environmental Research Letters* 10 (2): 024009.

Jaffe, Adam B., Richard G. Newell, and Robert N. Stavins. 2005. "A Tale of Two Market Failures: Technology and Environmental Policy." *Ecological Economics* 54 (2–3): 164–74.

Jakob, Michael, Stavros Afionis, Max Åhman, Angelo Antoci, Marlene Arens, Fernando Ascensão, Harro van Asselt et al. 2022. "How Trade Policy Can Support the Climate Agenda." *Science* 376 (6600): 1401–03.

Jessop, Simon, Seham Eloraby, and Valerie Volcovici. 2022. "IMF Chief Says $75/Ton Carbon Price Needed by 2030." *Reuters*, November 7. https://www.reuters.com/business/cop/exclusive-cop27-imf-chief-says-75ton-carbon-price-needed-by-2030-2022-11-07/.

Jiang, Betty, Maheep Mandloi, Richard Carlson, Matt Hope, Mike Ziffer, Nick Campanella, Ariel Rosa et al. 2022. *Treeprint: U.S. Inflation Reduction Act–A Tipping Point in Climate Action*. Zurich: Credit Suisse.

Kaufman, Noah. 2023. "Should the Inflation Reduction Act End the Dream of a National Carbon Price?" Center for Global Energy Policy, January 26. https://www.energypolicy.columbia.edu/should-the-inflation-reduction-act-end-the-dream-of-a-national-carbon-price/.

Kleimann, David, Niclas Poitiers, André Sapir, Simone Tagliapietra, Nicolas Véron, Reinhilde Veugelers, and Jeromin Zettelmeyer. 2023. "How Europe Should Answer the US Inflation Reduction Act." Bruegel Policy Contribution 04/23.

Le Maire, Bruno, and Robert Habeck. 2022. *For a European Green Industrial Policy*. Paris and Berlin: Ministry of the Economy, Finance and Industrial and Digital Sovereignty of France and Ministry of Economic Affairs and Climate Action of German.

Liu, Zhu, Jing Meng, Zhu Deng, Ping Lu, Dabo Guan, Qiang Zhang, Kebin He, and Peng Gong. 2020. "Embodied Carbon Emissions in China-US Trade." *Science China Earth Sciences* 63 (10): 1577–86.

Nemet, Gregory F. 2019. *How Solar Energy Became Cheap: A Model for Low-Carbon Innovation*. New York: Routledge.

Nordhaus, William. 2015. "Climate Clubs: Overcoming Free-Riding in International Climate Policy." *American Economic Review* 105 (4): 1339–70.

OECD. 2021. *Effective Carbon Rates 2021: Pricing Carbon Emissions through Taxes and Emissions Trading*. Paris: OECD Publishing.

OECD. 2022. *Pricing Greenhouse Gas Emissions: Turning Climate Targets into Climate Action*. Paris: OECD Publishing.

Parker, George, Andy Bounds, and Aime Williams. 2022. "Britain Joins EU in Criticising Biden's Green Subsidies Package." *Financial Times*, December 22. https://www.ft.com/content/21fe9241-5e1b-409e-8d6e-98f2e4d06793?emailId=b30d2e5a-5ffa-4653-8fec-c7b3fbb68a45&segmentId=13b7e341-ed02-2b53-e8c0-d9cb59be8b3b.

Parry, Ian. 2021. "Five Things to Know about Carbon Pricing." *IMF Finance and Development*, September. https://www.imf.org/en/Publications/fandd/issues/2021/09/five-things-to-know-about-carbon-pricing-parry.

Parry, Ian, Simon Black, and James Roaf. 2021. "Proposal for an International Carbon Price Floor among Large Emitters." IMF Staff Climate Notes 2021/001.

Penn-Wharton Budget Model. 2023. "Update: Budgetary Cost of Climate and Energy Provisions in the Inflation Reduction Act." April 27. https://budgetmodel.wharton.upenn.edu/estimates/2023/4/27/update-cost-climate-and-energy-inflation-reduction-act.

Rodrik, Dani. 2022. "Climate before Trade." *Project Syndicate*, December 13. https://www.project-syndicate.org/commentary/fighting-climate-change-priority-over-wto-trade-rules-by-dani-rodrik-2022-12?barrier=accesspaylog.

Rorke, Catrina, and Greg Bertelsen. 2020. *America's Carbon Advantage*. Washington, DC: Climate Leadership Council.

Roy, Nicholas, Dallas Burtraw, and Kevin Rennert. 2021. "Cost Analysis and Emissions Projections under Power Sector Proposals in Reconciliation." Resources for the Future Issue Brief 21-15.

Shapiro, Joseph S. 2021. "The Environmental Bias of Trade Policy." *Quarterly Journal of Economics* 136 (2): 831–86.

Singhania, Monica, and Neha Saini. 2021. "Demystifying Pollution Haven Hypothesis: Role of FDI." *Journal of Business Research* 123: 516–28.

Stangarone, Troy. 2022. "Inflation Reduction Act Roils South Korea-US Relations." *The Diplomat*, September 20. https://thediplomat.com/2022/09/inflation-reduction-act-roils-south-korea-us-relations/.

Swanson, Ana. 2022. "U.S. Proposes Green Steel Club That Would Levy Tariffs on Outliers." *New York Times*, December 7. https://www.nytimes.com/2022/12/07/business/economy/steel-tariffs-climate-change.html.

Timilsina, Govinda R. 2022. "Carbon Taxes." *Journal of Economic Literature* 60 (4): 1456–1502.

US Census Bureau. 2018–2020. "Annual Survey of Manufactures: Summary Statistics for Industry Groups and Industries in the U.S.: 2018–2021." United States Department of Commerce. https://data.census.gov/table?q=AM1831BASIC&t=Business+and+Economy:Expenses+and+Expenditures:Value+of+Sales,+Receipts,+Revenue,+or+Shipments&n=N0400.00:N0600.00&tid=ASMAREA2017.AM1831BASIC01 (accessed May 2023).

US Census Bureau. 2019. "Exports and Imports by NAICS Commodities." USA Trade Online. United States Department of Commerce. https://usatrade.census.gov/index.php (accessed May 2023).

US Department of the Treasury. 2023. *General Explanations of the Administration's Fiscal Year 2024 Revenue Proposals*. Washington, DC: U.S. Department of the Treasury.

Weise, Zia. 2021. "EU's Looming Carbon Tax Nudged Turkey toward Paris Climate Accord, Envoy Says." *Politico*, November 6. https://www.politico.eu/article/eu-carbon-border-adjustment-mechanism-turkey-paris-accord-climate-change/.

Wiedmann, Thomas, and Manfred Lenzen. 2018. "Environmental and Social Footprints of International Trade." *Nature Geoscience* 11: 314–21.

Wood, Richard, Michael Grubb, Annela Anger-Kraavi, Hector Pollitt, Ben Rizzo, Eva Alexandri, Konstantin Stadler, Dan Moran, Edgar Hertwich, and Arnold Tukker. 2020. "Beyond Peak Emission Transfers: Historical Impacts of Globalization and Future Impacts of Climate Policies on International Emission Transfers." *Climate Policy* 20 (S1): S14–27.

World Bank. 1990-2023. "Carbon Pricing Dashboard." https://carbonpricingdashboard.worldbank.org/ (accessed May 2023).

World Bank. 2014. "Statement: Putting a Price on Carbon." World Bank, June 3. https://www.worldbank.org/en/programs/pricing-carbon/statement.

World Bank. 2022. "What You Need to Know about Article 6 of the Paris Agreement." World Bank, May 17. https://www.worldbank.org/en/news/feature/2022/05/17/what-you-need-to-know-about-article-6-of-the-paris-agreement.

Zabanova, Yena. 2021. *Navigating Uncharted Waters: Russia's Evolving Reactions to the CBAM and the European Green Deal*. Cologne: Heinrich Böll Foundation.

Zhu, Yongbin, Yajuan Shi, Jing Wu, Leying Wu, and Wen Xiong. 2018. "Exploring the Characteristics of CO2 Emissions Embodied in International Trade and the Fair Share of Responsibility." *Ecological Economics* 146: 574–87.

Journal of Economic Perspectives—Volume 37, Number 3—Summer 2023—Pages 163–188

Global Transportation Decarbonization

David Rapson ⓡ Erich Muehlegger

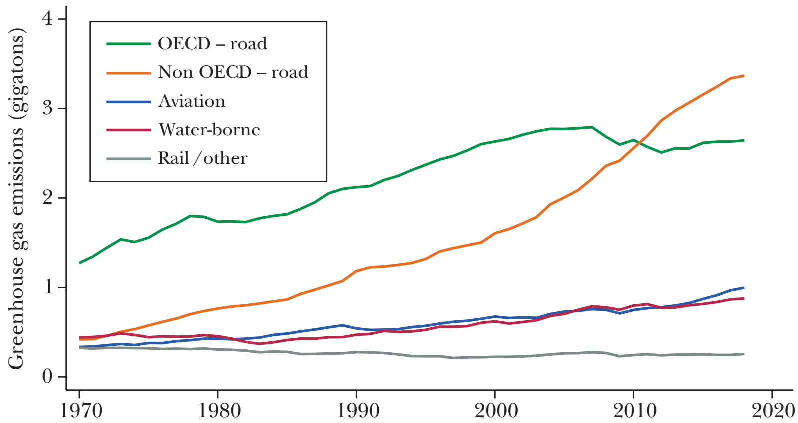
The benefits of the transportation sector outweigh its environmental costs by orders of magnitude. For instance, transportation is a prerequisite to international trade, and despite generating roughly 2.4 gigatons of CO_2 emissions annually—just under 7 percent of total global emissions from fossil fuels and industry—through the geographic redistribution of goods, Shapiro (2016) estimates that gains from international trade outweigh emissions-related climate damages by a factor of 161-to-1. In addition, transportation facilitates the movement of people within and across urban areas, creating benefits for workers and firms and generating distributional benefits for low-income and disadvantaged households by alleviating spatial mismatches between supply and demand in labor markets.

Such dramatic differentials in costs and benefits highlight the profound tradeoffs confronted by emissions abatement efforts in the transportation sector. Decarbonization must be implemented in a manner that supports the continued provision of low-cost transportation services, or risk eroding the foundation of the local and global economies.

■ *David Rapson is Chancellor's Leadership Professor and Erich Muehlegger is Professor of Economics, both at University of California—Davis, Davis, California. Rapson is also a Policy Advisor and Senior Research Economist, Federal Reserve Bank of Dallas, Dallas, Texas. Muehlegger is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are dsrapson@ucdavis.edu and emuehlegger@ucdavis.edu. Their emails are dsrapson@ucdavis.edu and emuehlegger@ucdavis.edu. The order in which the authors names appear has been randomized using the AEA Author Randomization Tool (E6wTwGwaIvec), denoted by ⓡ.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.3.163.

Figure 1
Transportation Emissions by Sector



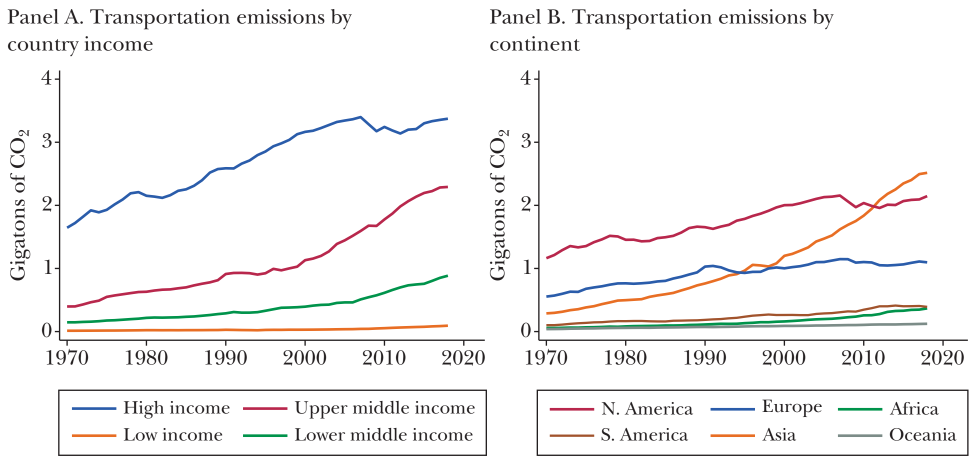
Source: Emissions Database for Global Atmospheric Research (European Commission 2023).
Note: This figure plots annual emissions greenhouse gas emissions (in gigatons) for five transportation sectors from 1970 to 2018.

At present, the vast majority of transportation services rely on fossil fuels as the primary source of propulsion energy. Nearly 100 million barrels per day of crude oil are processed primarily into gasoline, diesel, and jet fuel for transportation. Emissions from transportation have increased at roughly 2 percent per annum for the past five decades and are closely linked to economic growth. Over a similar time frame, transportation's share of total greenhouse gas emissions has risen from roughly 18 to 21 percent (based on authors' calculations from European Commission 2023). As noted by the Fifth Assessment of the Intergovernmental Panel on Climate Change, transportation emissions are likely to continue to increase by roughly 50 percent over the next 30 years in the absence of substantial carbon mitigation (Intergovernmental Panel on Climate Change 2014).

Four sectors account for over 97 percent of global greenhouse gas emissions from transportation: (1) on-road transportation in developed (OECD) countries (32.4 percent); (2) on-road transportation in developing (non-OECD) countries (41.4 percent); (3) maritime shipping (10.8 percent); and (4) air transportation (12.2 percent). Rail and other forms of transportation are comparatively negligible contributors to global emissions.

In Figure 1, we plot the evolution of global greenhouse gas emission estimates from these subsectors from 1970 to 2018, based on European Commission (2023). For comparison, worldwide greenhouse gas emissions across all sectors of the economy were roughly 36 gigatons in 2018 (IEA 2022b). Figure 1 suggests two themes that will recur throughout the essay: the centrality of road vehicles in the task of decarbonizing transportation and the ongoing rise in transportation

Figure 2
Transportation Emissions



Source: Emissions Database for Global Atmospheric Research (European Commission 2023).
Note: This figure plots annual emissions greenhouse gas emissions (in gigatons) for five transportation sectors from 1970 to 2018.

emissions in developing countries. In 2018, on-road transportation accounted for roughly three-quarters of transportation emissions. The patterns of road emissions in the higher-income countries in the OECD peaked in 2008 and have maintained a slightly lower level and flat trajectory in recent years. For the first time in 50 years, road emissions in these higher-income countries appear to have become unlinked from economic growth. Total road emissions in other non-OECD countries, on the other hand, have overtaken OECD emissions and continue to grow. Likewise, emission from maritime shipping and air transport have risen consistently over the past five decades. Emissions from maritime shipping and air transport grew by 1.5 percent and 2.3 percent per annum between 1970 and 2018 and now account for roughly 23 percent of transportation greenhouse gas emissions.

The trajectories of emissions in OECD and non-OECD countries are consistent with the predictions of the environmental Kuznets curve, a concept introduced by Grossman and Krueger (1991) and discussed in this journal by Dasgupta et al. (2002), who suggest that countries in the process of economic development see a sharp rise in environmental costs for a time, later followed by a leveling off and decline. The left panel of Figure 2 disaggregates emissions by GDP quartile. High-income country emissions mirror the OECD plateau described above. Upper- and lower-middle income countries are in high- and low-growth phases, respectively, while low-income countries exhibit low demand for transportation services. As economic development proceeds, demand for transportation services grows. This is particularly clear in upper-middle and lower-middle income countries in Asia, where emissions have risen ninefold since 1970 (a rate of roughly 4.7 percent per annum over half a century).

Asia, the most populous continent, has experienced rapid economic growth in recent decades and is now the largest contributor to transportation emissions (as seen in the right panel of Figure 2, which disaggregates emissions by geography). In coming decades, Africa will almost surely emerge as an important contributor to transportation emissions growth. Since the 1980s, sub-Saharan Africa has experienced the fastest population growth of any region in the world. It is expected to add nearly one billion people by 2050, nearly doubling its population (United Nations Department of Economic and Social Affairs 2022). While predictions of per capita income growth in the decades ahead are inevitably uncertain (World Bank 2022), aggregate demand for transportation services will nonetheless increase dramatically in coming decades due to population growth alone.

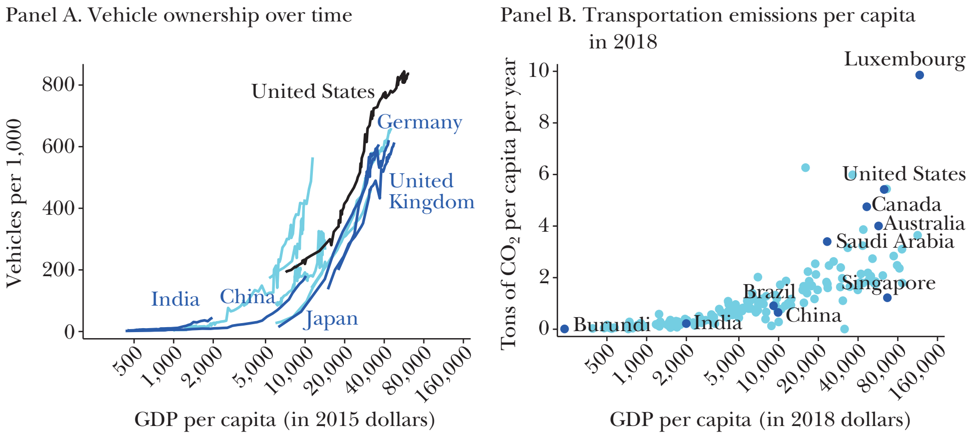
Per capita income growth will only contribute to the growth in emissions, as the demand for transportation is strongly correlated with per capita income. This can be seen most readily in historical patterns of vehicle ownership. The left panel of Figure 3 traces the path of vehicle ownership and per capita GDP over time in the United States, Germany, Japan, and the United Kingdom. The trajectories for India and China over the same time appear in the bottom left of the figure. The expansion of vehicles is a substantial driver of the strongly positive relationship between per capita GDP and per capita carbon emissions from transportation, shown in the right-hand panel of Figure 3. If China and India (and other developing countries) follow the pattern of today's developed economies, they are in the early stages of a prolonged period of rapidly accelerating demand for transportation services. As Intergovernmental Panel on Climate Change (2014) wrote:

> Without aggressive and sustained mitigation policies being implemented, transport emissions could increase at a faster rate than emissions from the other energy end-use sectors and reach around 12 Gt CO_2eq/yr by 2050. Transport demand per capita in developing and emerging economies is far lower than in Organisation for Economic Co-operation and Development (OECD) countries but is expected to increase at a much faster rate in the next decades due to rising incomes and development of infrastructure.

Electrification: Advantages and Limitations

Replacing fossil fuels in the name of decarbonization is necessary but will be particularly difficult due to their as-yet unrivaled bundle of attributes: abundance, ubiquity, energy density, transportability, and cost (Covert, Greenstone, and Knittel 2016). Yet, in the developed world, there is a growing commitment to electrification as the dominant pathway to a meaningful reduction in road transportation emissions. One of the appeals of the electrification vision is that much of the technology already exists at commercial scale, and costs have been declining

Figure 3
Transportation Demand and Income



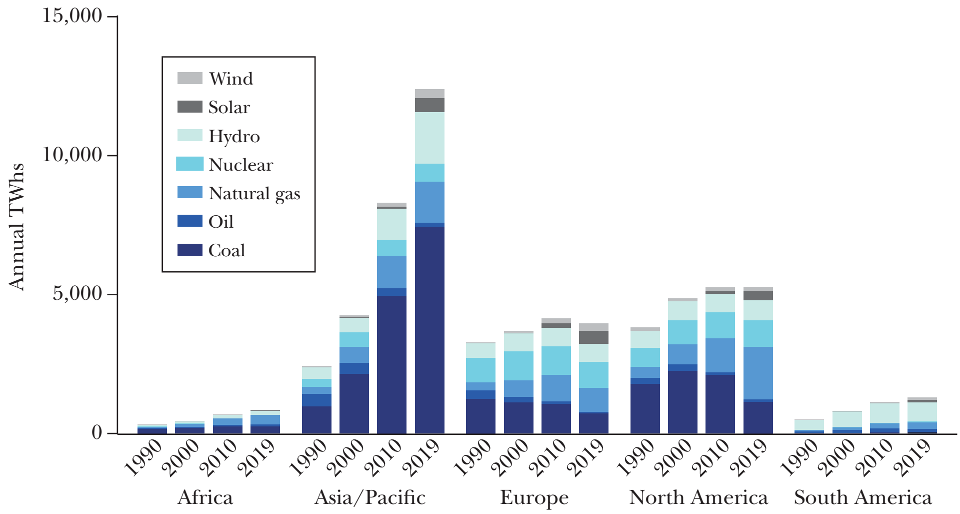
Source: Vehicles per capita (Davis and Boundy 2022); Transportation Emissions (European Commission 2023); GDP per capita (World Bank 2023).
Note: The left panel plots vehicles per capita against real GDP per capita for major economies over time. Country series begin in 1900 for the United States, 1960 for Canada, France, Germany, Japan, and the United Kingdom, 1980 for China, and 1985 for Argentina, Brazil, India, Indonesia, Korea, Malaysia, and Pakistan. The right hand panel plots emissions per capita against real GDP per capita in 2018, with major economies and outliers highlighted.

steeply in recent years. The approach favored by policymakers in developed countries is to shift simultaneously towards greener sources of electricity generation while promoting adoption of electric vehicles in an attempt to reduce their cost. Although obstacles exist, there are reasons for optimism about this path.

Electric vehicles are getting cheaper. This is driven mainly by reductions in battery costs, which fell by 14 percent per annum from 2007 to 2014 (Nykvist and Nilsson 2015) and have continued to decline since. Over the past decade, the speed at which battery costs declined exceeded even the most optimistic of earlier projections (as discussed in Knittel 2012). Many expect electric vehicles to achieve price parity with gasoline powered vehicles within the next decade (National Academies of Sciences, Engineering, and Medicine 2021). An expanding slate of electric light-duty vehicle models is being sold, targeting different price points and a broader set of consumer preferences.

The grid is getting cleaner. In Europe and North America over the past 20 years, the electric grid has shifted towards less carbon-intensive sources of power in both cases (Figure 4). In North America, natural gas has displaced coal as the dominant source of electricity and the grid has absorbed substantial growth of wind and solar power. On the margin, electric vehicles now generate unambiguously lower greenhouse gas externalities than gasoline-powered vehicles (Holland et al. 2020) wherever coal is not the marginal source of electricity (so in most of the country).

Figure 4
Electricity Generation Mix over Time, by Region



Source: IEA (2022a).
Notes: The bars reflect terawatt-hours of electricity generation by region and fuel source for 1990, 2000, 2010, and 2019. As a rough point of reference, one terawatt-hour of electricity is enough power to light over a million homes for one year or cool half a million homes for a year.

Renewable energy comprised over 20 percent of electricity generation in 2021, double its contribution from a decade earlier. In Europe, over the past two decades solar and wind generation has grown rapidly, replacing coal on a one-for-one basis.

Governments are directing the full strength of their conviction behind electrification. The electric vehicle market share (of new sales) has grown to over 14 percent globally in 2022, driven by enthusiastic early-adopters, large government incentive programs, and the aforementioned 90 percent decline in battery costs (IEA 2023). Policymakers extrapolating early successes into the future appear to conclude that electric vehicles will render gasoline cars obsolete within two decades. As of this writing, the European Union, China, Japan, South Korea, several US states, and many others have declared their intention to ban gasoline and diesel cars. The force and magnitude of these efforts are, in effect, choosing electrification as *the* winner of the decarbonization sweepstakes in rich countries.

However, it would be risky to extrapolate from recent trends what the world may look like in the future. There is no guarantee that the electric grid will remain reliable as we replace the most flexible sources of supply with intermittent renewables. There is no guarantee that batteries, which require enormous quantities of increasingly scarce metals, will continue to enjoy steady cost declines. And there is no guarantee that the political will to support electrification will continue if cost and reliability concerns become reality.

In our view, electrification is the most likely technology pathway for deep transportation decarbonization. Yet there are reasons to be skeptical of the aspirations for a fully electric transportation future. This skepticism applies to both the rich and developing worlds. Rapson and Bushnell (forthcoming) offer a discussion of the limitations of electric vehicles even in the rich world, where the electric grid is advanced and resources are relatively abundant. In what follows, we take a global perspective to describe several obstacles that will need to be overcome for electrification to become the default transportation energy source for light-duty road transportation.

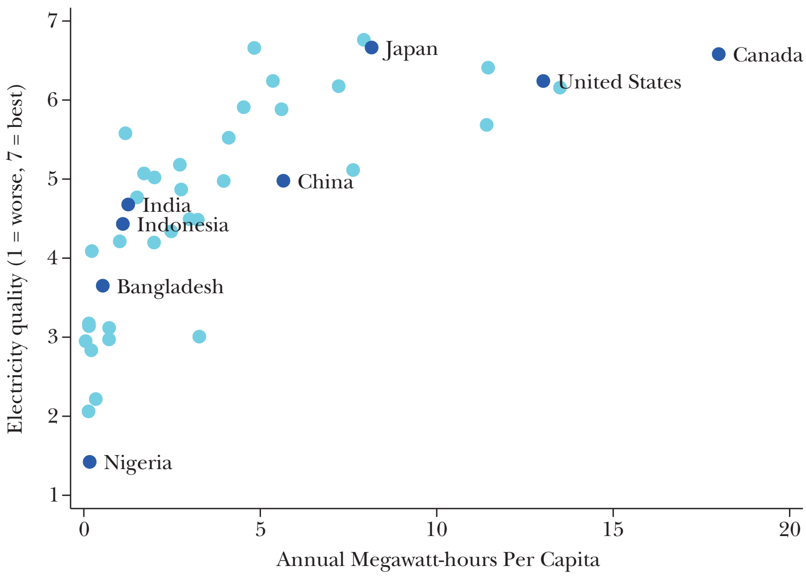
Electricity Reliability in the Developing World

The electrification vision requires squinting, or even a dream-like state, when considering its prospects in the developing world, where fossil fuels dominate electricity generation. China is a revealing case study. It is on track to put more electric vehicles on the road this year than the rest of the world combined (Wakabayashi and Fu 2022). However, the environmental benefits of this shift are more modest because China's investments in electricity generation capacity and grid infrastructure over the past several decades have been dominated by coal (Qiao et al. 2019). In Asia overall, new coal generation outstripped new "renewable" generation by a factor of *five* over 2000–2019 (as shown in Figure 4).[1]

Many developing countries also face the hurdle of improving electricity distribution infrastructure and grid reliability. Figure 5 plots country-level generation per capita (on the x-axis) and a proxy for the reliability of electricity (on the y-axis), the average response by country business leaders to the World Economic Forum Global Competitiveness Report survey question, "In your country, how reliable is the electricity supply (lack of interruptions and lack of voltage fluctuations)? [1 = extremely unreliable; 7 = extremely reliable]." Electrification of transportation requires both sufficient generation capacity and a reliable grid. Wealthy nations score highly on electricity availability and reliability. But most developing countries have less reliable electricity as well as substantially lower levels of generation per capita. Distributed solar microgrids are unlikely to substitute perfectly for a centralized grid (Lee, Miguel, and Wolfram 2016). Moreover, the scale of incremental fixed investment required for widespread electrification might prove prohibitive for many developing countries and may first require addressing other market failures impeding electricity infrastructure investment, such as imperfect contract enforcement (Ryan 2020), and insufficient regulated tariffs (Blimpo, McRae, and Steinbuks 2018). For a back-of-the-envelope estimate, 4,000 miles per capita of annual travel requires roughly 1 megawatt-hour of electricity per capita, each year. Even with rapid development, Chinese generation per capita only rose 4.5 megawatt-hours per capita per annum over the past three decades. Moreover, while vehicle electrification

[1] Admittedly, much of this increase is the result of the expansion of China's electricity industry: roughly 60 percent of generation and 67 percent of coal generation in Asia occurred in China in 2020. But coal generation also grew substantially in Asia outside of China, more than quadrupling from 1990 to 2020.

Figure 5
Per Capita Generation and Electricity Reliability



Sources: Electricity Reliability (World Bank 2019); Electricity Generation (IEA 2022a).
Notes: This figure plots electricity quality and annual electricity generation capita by country in 2018. Electricity quality is measured a scale of 1 to 7 and reflects the average response by business leaders to the survey question to the World Economic Forum, Global Competitiveness Report: "In your country, how reliable is the electricity supply (lack of interruptions and lack of voltage fluctuations)? [1 = extremely unreliable; 7 = extremely reliable]." Generation is measured in megawatt-hours per capita. Select countries are highlighted.

is one possible electricity end use, transportation would compete directly with other uses of additional electricity with high marginal value to households (for example, Dinkelman 2011) and firms (for example, Allcott, Collard-Wexler, and O'Connell 2016), including lighting, cooling, and powering industrial equipment.

High Costs of Electrification

Even in rich countries, there are reasons to expect the marginal costs of electrification to rise, not fall, as the share of electric vehicles increases (Rapson and Bushnell forthcoming). To date, demand for electric vehicles in the United States has been concentrated among wealthy, highly-educated buyers who express concern about climate change (Davis 2019; Archsmith, Muehlegger, and Rapson 2021). These buyers tend to own multiple cars and live in single-family homes in coastal states or the suburbs of large cities. To achieve full (or even deep) electrification, adoption of electric vehicles will need to extend into new consumer segments. Two of these include low- and middle-income households who are interested in adopting an electric vehicle, and rural Americans who tend to prefer light trucks to sedans and are less compelled to make decisions based on concerns about climate change.

A multitude of practical obstacles to adoption of electric vehicles arise for these customer segments. Lower-income households tend to have smaller vehicle portfolios, and thus cannot easily hedge their transportation needs across different vehicle types. For these buyers, electric vehicles are a more expensive and potentially less reliable alternative to gasoline cars. Many of these potential buyers live in multi-unit buildings that tend not to offer on-site charging options. Rural consumers tend to prefer larger vehicles, which are currently not widely available in an electric drivetrain. While new models of electric vehicles are already being introduced to meet some of these needs, it remains to be seen how popular they will be among this sub-population. Finally, physical obstacles exist even beyond the well-known multi-unit dwelling issue. Rapson and Bushnell (forthcoming) estimate that roughly 20 percent of US single family homes would require an electric system upgrade in order to accommodate a dedicated "level 2" charger.

Public costs of electric vehicle adoption are already high and are likely to increase. Despite progress, the carbon intensity of the electric grid remains a challenge, even in developed countries. Almost 60 percent of US electricity was generated from coal (21 percent) and natural gas (36 percent) in 2022 (EIA 2023). Substituting towards more solar and wind energy is inexpensive from an energy production perspective, but must be supported by transmission (long-haul) and distribution system ("last mile") infrastructure to transport energy to consumers. Such upgrades range from costly to potentially impossible. Elmallah, Brockway, and Callaway (2022) estimate that distribution system upgrades in California will cost between $200 and $2,000 per household, depending on the ability of electric utilities to shift the timing and location of demand on the grid. Davis, Hausman, and Rose (2023) paint an even more discouraging picture about the prospects for transmission investments, the amount of which needs to triple in order to integrate sufficient clean electricity to achieve net-zero goals by 2050 (Pascale, Jenkins, and Leslie 2021). Such investments encounter obstacles relating to permitting, the current process for which is distributed in a manner that gives property owners on the right-of-way a string of potential vetoes.

The Battery Supply Chain

Demand for electric vehicle batteries doubled in 2021, and prices for key battery inputs rose by as much or more. The price of lithium (an ingredient to all electric vehicle battery chemistries in use today) was recently seven times greater than at its 2020 trough, though it has since fallen. Prices of both nickel and cobalt doubled over a similar timespan. A dramatic expansion of the battery supply chain will be necessary to meet demand under existing transportation electrification policies, with an even larger expansion required to meet stated future goals. The IEA (2022c) estimates that global battery anode and cathode production will be required to expand by six to ten times present day volumes to meet 2030 demand under these scenarios.

Such a dramatic expansion of battery production requires unprecedented growth to occur at each link of a complicated battery supply chain. The supply chain has three main links, or levels. "Upstream," raw materials for production must be extracted. Precisely which battery minerals are needed depends on the battery chemistry, which is an endogenous choice made by automakers and battery manufacturers (we will come back to this). In the "midstream" segment, raw minerals are processed and intermediate battery components (cathodes and anodes) are produced. Finally, battery cells are produced and linked in "packs" that can be used in electric vehicles, which is referred to as the "downstream" segment.

Expanding production in each of these links on the chain requires long lead times. According to IEA (2022c), developing new lithium and nickel extraction sites can take between 5 and 20 years; raw materials processing and cathode/anode production facilities requires between 2 and 8 years; and battery production facilities between 1 and 5 years. In this section we assess the prospects for success, and describe a wide array of complexities and costs associated with the task ahead.

The Good. While the required supply chain expansion is enormous, there are reasons to be optimistic that we can make substantial progress in the next 10–20 years. Primary among these is evidence that governments and industry participants are already responding to economic incentives. When confronted with high nickel and cobalt prices, for example, China and Tesla have shifted towards alternate battery chemistries. Lithium-iron-phosphate batteries sacrifice some energy density relative to others, but eliminate the need for nickel, cobalt, and magnesium entirely (though they do nothing to reduce demand for lithium). Half of Teslas produced in 2022 will use these batteries. China had already prioritized lithium-iron-phosphate batteries to take advantage of patent expirations, and because their focus on shorter-range cars in the domestic market makes these batteries more suitable. These decisions will relieve pressure on some of the upstream bottlenecks, at least in the short run. High mineral prices will also stimulate supply expansions. Policymakers and private firms alike are aware of the need to expand midstream and downstream capacity, and abundant capital is flowing towards these areas.

The Bad. Due to long lead times required to expand at any level of the supply chain, the status quo exhibits strong inertia. This is particularly concerning to Western countries who currently rely on China and Russia for key inputs. Russia dominates the market for battery grade nickel, and China dominates midstream and downstream activities across the board. IEA (2022c) reports that over half of global capacity for lithium (~60 percent), cobalt (65 percent) and graphite (70 percent) processing resides in China. China has an even larger share of cell component production (70–85 percent) and battery cell production (75 percent). Many have expressed concerns about relying on China for critical inputs in this time of geopolitical adversity.

How big a problem is this for the West? Our view is that it is less problematic than one might think. A strategic Chinese battery supply disruption would harm China economically and is unlikely to produce the jarring economic adjustments caused by a major OPEC supply disruption in global oil markets or the Russian suspension of natural gas exports to Europe. Still, the strategic, if not economic, benefits to diversifying and even onshoring some midstream and downstream capabilities are hard to predict and potentially substantial.

Relieving supply chain bottlenecks through reuse and recycling of batteries is, at present, unlikely to provide a solution. Few batteries for electric vehicles are in circulation today relative to future demand, and the profit margins in recycling are typically not high. IEA (2022c) estimates that less than 1 percent of 2030 lithium and nickel demand will be met from recycling. Cobalt is only slightly better, at under 2 percent.

The Ugly. A shift to electric vehicles, at least to some degree, amounts to trading greenhouse gas reduction benefits for local environmental and social damages relating to the battery supply chain (Lee et al. 2020). While the electric vehicle transition may nonetheless pass a global cost-benefit test in the long run, the (often) severe environmental and social costs to local communities supplying the minerals cannot be ignored. The most notorious and in our view most objectionable instance is cobalt mining in the Democratic Republic of Congo. The Democratic Republic of Congo produces the majority of global cobalt supply and has a reputation for using unsafe mining practices and child labor (Kara 2023). In Chile, mining for lithium has disrupted local ecosystems due to the use of evaporation pools created by converting meadows and lagoons into salt flats, a process that has depleted groundwater across the Atacama Desert (Lee et al. 2020). A promising source of abundant reserves of lithium, cobalt, magnesium, and nickel exists at the bottom of the Pacific Ocean, but it is difficult to envision how it can be made accessible without destroying substantial (multiple square miles) of the ocean floor. Our ability to assign an accurate value to these nonmarket goods is poor, yet the moral, social, and ecological stakes are high.

To summarize, producing enough electric vehicle batteries to meet demand through 2030 is possible, but will be costly and requires careful planning and patience. Supply chain constraints may directly influence the cost and desirability of electric vehicles. For example, earlier we lauded Tesla's decision to use lithium-iron-phosphate batteries as a way to relax contemporary upstream constraints, but it does so at the cost of electric vehicle range. Average battery size increased by 60 percent between 2015 and 2021 (IEA 2022c). While many electric vehicle drivers likely do not need a 300-mile range battery, one of the main industry and policy goals in recent years has been to overcome range anxiety, which is seen as an obstacle to widespread adoption, particularly for high-use drivers or drivers living in cold areas where range declines. Innovation may help, but likely only in the medium- and long-run. IEA (2022c) mentions two promising technologies in the upstream segment. Direct lithium extraction bypasses the need to evaporate unconcentrated brine. If successful, this will drive down costs and lead-times for capacity expansion, as well as dramatically reducing local environmental damages. It is being piloted today. Reliance on Russian battery-grade "class 1" nickel could be reduced by producing class 1 nickel from class 2 nickel, for which Australia is the world's largest supplier. However, this process is twice as capital-intensive, takes longer, and is three times as carbon-intensive as present class 1 nickel mining methods. Early-stage deployments have encountered cost overruns and project delays.

It is quite possible that you, our reader, may read some of these “underappreciated challenges” and wonder in what world these are not obvious. But for each of those such readers, we suspect there is another kind who views emphasizing these challenges as unnecessary dithering about minor details that ought to be subservient to saving the planet. To this, we can only emphasize that our view arises from acknowledging that tradeoffs exist. The stakes are high from all perspectives. If renewable electricity and electric vehicles were superior to fossil fuels and the internal combustion engine in every dimension, there would be little need to write this paper. Our goal is to highlight costs of electrification that we view as nontrivial and worthy of serious consideration by climate and energy policymakers as they weigh the costs and benefits of various paths forward.

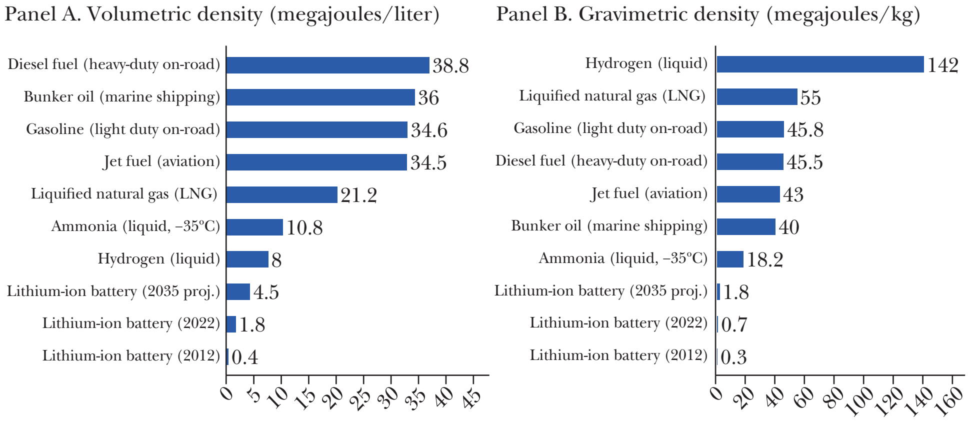
What Alternatives Exist to Decarbonize Other Sectors?

Electrification is unlikely to be a viable technology pathway for transportation segments that require very large amounts of energy and/or have extreme energy density requirements. The primary alternatives to electrification are “renewable” transportation fuels. These include: biofuels, chemically-similar substitutes for petroleum-based transportation fuels (gasoline and diesel) produced from biomass; hydrogen, which can be combined with oxygen in a fuel cell to produce energy and water vapor; and other alternatives, such as liquified natural gas or ammonia. Although these fuels take different forms, they share three potential advantages that would bypass the dramatic expansion that would otherwise be needed for the electric grid in developing countries and offer a viable long-run alternative for maritime trade and air travel.

First, renewable transportation fuels can be (and are) transported over long distances. They potentially provide a trade-based decarbonization pathway for developing countries, rather than electrification that requires local generation and distribution infrastructure and faces the hurdles and road-blocks already described.

Second, some (though not all) renewable transportation fuels can “drop into” existing fuel supply chains and engine, further reducing upfront investment or switching costs relative to electrification. Biofuels, which are refined to be interchangeable with gasoline and diesel fuel, offer a particular advantage here. They are designed to be roughly equivalent, can be blended to different degrees with existing fuel depending on use, used in conventional engines, and transported, stored, and distributed through similar infrastructure. US drivers are already familiar with ethanol, one of the most well-known biofuels, which in the United States is blended with gasoline and comprises between 10 and 15 percent of each gallon of “gasoline” purchased at the retail pump. Ethanol-blending, at these levels, offers similar performance to pure gasoline, but does lower the overall energy density, reducing fuel economy by about 3 percent per gallon of fuel. In the near term, sustainable

Figure 6
Energy Density of Transportation Fuels



Sources: Allain et al. (2012); Schlachter (2012); National Academices of Sciences, Engineering, and Medicine (2016); Tran, Palomino, and Oliver (2018); Kargul et al. (2019); Holladay, Abdullah, and Heynu (2020); Rodrigue (2020); Esau (2021); Statista (2020); US Department of Energy (2021, 2023b).
Note: Values for volumetric and gravimetric density are collected from a range of sources and compiled by the authors.

aviation fuel offers the best opportunity for reducing carbon-intensity in aviation, as it is operationally indistinguishable from jet fuel, offers similar energy density, and does not require any additional investment or regulatory approval to be blended at a 1:1 ratio with jet fuel.

Finally, these fuels are more "energy dense" than electric batteries, storing more energy per unit of space (volumetric energy density) or in a given amount of weight (gravimetric energy density), and providing a pathway for sectors for which electrification is unlikely to offer a solution (at least over the timeframe of the next several decades). Whether evaluated on a volume basis or a weight basis (the left and right panels of Figure 6, respectively), petroleum-based and alternative liquid fuels offer energy densities that are one and sometimes two orders of magnitude greater than current (and projected) lithium ion batteries. These constraints are particularly relevant for air transport where both space and weight for fuel are paramount considerations for any lower-carbon aviation fuel alternatives, but also relevant for ocean-borne freight that traverses long-distances without refueling. In both cases, electrification is unlikely to meet the industry needs in the near term. As one example, fully battery-reliant systems for commercial air travel are viewed as unlikely to develop beyond small private aircraft for the next few decades. Anticipated energy requirements for sustained, even short-distance commercial flights would require battery with energy density greater than 6.5 megajoule per kilogram, relative to a projection of 1.8 megajoule per kilogram by 2035 (National Academies of Sciences, Engineering, and Medicine 2021).

Drawbacks of Alternative Fuels

The potential for cost-effective carbon mitigation depends on the how the alternative fuel is produced. The carbon benefits and potential quantity of different biofuels depend largely on the biomass "feedstock" used and on the efficiency of the refining process (EPA 2023). First-generation biofuels are produced from consumable "feedstock," like corn, sugar cane, and oilseed crops. These feedstocks are the most straightforward to process and account for the majority of current biofuel production. Corn- and cane-based ethanol are both cost competitive with gasoline at roughly \$50–\$75 per barrel of oil; first-generation biodiesel is cost competitive at \$80–\$120 per barrel (IEA 2022d). Although costs have fallen over time, the cost of sustainable aviation fuel remains two to three times more expensive than petroleum-based jet fuel (Congressional Research Service 2022). The price premium associated with sustainable aviation fuel has limited its adoption amongst cost-conscious airlines—estimates place 2021 sustainable aviation fuel production at approximately 25 million gallons, relative to 13.7 *billion* gallons of jet fuel consumed by US airlines.

The carbon benefits of biofuels and sustainable avaiation fuel are undermined by the fuel and fertilizer used for cultivation of crops (Melillo et al. 2009) and by indirect shifts in the use of land for cultivation (Keeney and Hertel 2009; Searchinger et al. 2008). Although hydrogen is an alternative to electrification (and biofuels) and offers emission-free combustion, the carbon benefits and costs depend on the method by which hydrogen is produced. Presently, most hydrogen is produced by processing natural gas (Nikolaidis and Poullikkas 2017), which is substantially lower cost than carbon-free "green" hydrogen, produced by separating water into hydrogen and oxygen using solar or wind-based electrolysis. Similarly, ammonia as presently produced is both energy and carbon-intensive, accounting for roughly 2 percent of total worldwide energy consumption and generating roughly half a gigaton of carbon per year (IEA 2021).

All of the alternative fuels have opportunity costs that are particularly salient to policymakers. Biofuel feedstocks are also part of the food supply chain, placing energy end uses in direct competition with food. At present, roughly 15–20 percent of cereal production is used for biofuels (IEA 2022f). Roberts and Schlenker (2013) finds evidence that feedstock demand of commodities has a meaningful impact on commodity prices. Using roughly one-third of corn as ethanol feedstock (as in the United States) increases corn prices by roughly 20 percent. The estimated impact on crop prices are roughly comparable to those from Condon, Klemick, and Wolverton (2015), who conduct a meta-analysis of estimates from the food-versus-fuel debate literature. The direct competition for consumable resources and the modest carbon reduction benefits of first-generation biofuels (Hill et al. 2006) have motivate research into "second-generation" biofuels that rely on nonfood feedstocks, which include used cooking oil, switchgrass, and plant cellulose, and even "third-generation" biofuels that rely on cultivated algae as feedstock. Second-generation biofuels offer potential for higher carbon savings when they are not directly cultivated or are waste by-products (Havlík et al. 2011). Some of the these feedstocks offer the potential for development at scale on marginal cropland, avoiding direct competition with convention crops (Cai, Zhang, and Wang 2011). However, with the exception of used cooking oil, these biofuels are not cost competitive at current oil prices (Witcover and Williams 2020). Similarly, "third-generation" biofuels are not cost competitive, impose substantial demands on water supplies, and have not yet reached commercial scale. Likewise, ammonia's current use is as a fertilizer, and hydrogen (and natural gas) are key inputs into fertilizer production. In the future, demand for these products as transportation fuels may compete with traditional agricultural and industrial uses.

Finally, many technologies that cannot drop in to existing supply chains or leverage existing combustion technology face a similar "chicken-and-egg" problem as vehicle electrification. As one example, widespread use of hydrogen would requires a new transportation, storage, and delivery network, development of which has been impeded by high costs on both sides of this two-sided market. At present, hydrogen cars and fueling infrastructure are not economically competitive. To date, 54 hydrogen stations are open nationwide, all but one of which are located in California where large subsidies are available (US Department of Energy 2023a).

Energy Efficiency: Once More Unto the Breach

The absence of viable alternatives to liquid hydrocarbons for jet propulsion and maritime shipping highlights the value of getting more from less, where possible. So despite a checkered past when it comes to delivering energy savings,[2] energy efficiency makes it once again onto a list of desirable decarbonization pathways.

Emissions are a function of both the fuel used and the efficiency with which that fuel is transformed into usable power. Although efficiency gains in some sectors are less flashy than electrification or novel transportation fuels, they will likely be needed to reduce carbon intensity over the long-term. In sectors in which fuel costs are a significant component of overall costs, commercial firms have a strong incentive to seek efficiency gains. In the airline industry, the desire to lower fuel costs that average ~15–20 percent of total airline costs (US Department of Transportation 2019) has contributed to steadily increasing efficiency over the past five decades. Airline fuel usage per seat mile has fallen by roughly 2 percent per annum since 1970, while engine efficiency alone rose at an average rate of roughly 7 percent per decade (National Academeics of Sciences, Engineering, and Medicine 2016). Fuel economy improvements through engine efficiency gains, airframe weight reductions, and aerodynamic improvements are anticipated to continue at a rate similar to historical levels for the next several decades. In the significantly longer term, further operational efficiency gains might be realized through alternative engine technologies, such as engines powered by electricity

[2] For an incomplete list, see Fowlie, Greenstone, and Wolfram (2018), Allcott and Greenstone (2017), Jacobsen and Van Benthem (2015), Jacobsen (2013), and Allcott and Greenstone (2012).

generated from liquid fuels (National Academies of Sciences, Engineering, and Medicine 2021). The question is how much of the efficiency gain leads to carbon reductions, as opposed to increases in demand for energy services (Knittel 2011; Gillingham, Rapson, and Wagner 2016).

Similarly, short-run options for alternative fuels in maritime shipping are severely constrained. According to the International Energy Agency, the most promising alternative (lower carbon) fuel options are ammonia, hydrogen, and biofuels, although liquified natural gas and electricity may also play a role (IEA 2022e). The International Maritime Organization is in the process of initiating demonstration projects to allow the industry to gain experience with various technology alternatives and ideally to bring down costs. But these are seeds that will only bear fruit in the long run.

In the near-term, maritime regulators have turned first to energy efficiency. The main regulatory body, the International Maritime Organization, recently mandated that ship operators meet Energy Efficiency Existing Index standards, with the goal of reducing carbon intensity from all ships by 40 percent by 2030 compared to 2008 (International Maritime Organization 2021). While some technology investment can help, the most common compliance mechanism will be for older ships simply to slow down. A 10 percent drop in cruising speeds will cut fuel usage by almost 30 percent, according to marine sector lender Danish Ship Finance. However, this is not without cost. A first-order effect will be to reduce the available industry tonnage capacity as the time to transport a given cargo on a given route will, on average, increase. Since the ability to expand the size of the shipping fleet is constrained in the short run (shipyards worldwide are already pre-booked to operate at capacity until 2026), there is a direct tradeoff between decarbonization efforts and the cost of the shipping services that form the backbone of international trade.

This will not be the first time we have sought to rely on energy efficiency for emissions reductions. It appears as an important "wedge" in most abatement forecasts and, until recently, has been a primary pillar of US climate policy. Corporate Average Fuel Economy and Greenhouse Gas Emissions Standards have governed the rate of emissions from the light duty vehicle fleet for decades. Still, gasoline demand grew until its plateau in the mid-2000s, muddying the causal link between the policy and the intended outcome. The risk with efficiency-based standards is that compliance can be achieved without reducing aggregate energy use (for example, Holland, Hughes, and Knittel 2009). The Environmental Protection Agency is considering applying ever more stringent standards to the light duty car fleet. The question is whether these regulatory agencies can achieve the decarbonization goals by leaning on a policy that transmits weak incentives to market participants. In the case of both cars and maritime shipping, the compliance costs may be sufficiently large as to reduce the aggregate level of transportation services enjoyed in the economy. The economic costs could far outweigh the environmental benefits, even when approximated by using the most aggressive estimates of the social cost of carbon.

Implications for Policy

Decarbonizing transportation is a challenge of immense scope. It entails a transformation of how we move people and goods throughout the economy. Many countries are proceeding with aggressive policies that seek to speed this transition. Four challenges are likely be important in determining the success of the transition path.

Decoupling of Emissions and Income

Successful decarbonization involves the decoupling of transportation emissions from income. For developed economies, this step means reducing emissions from current levels; for developing countries, it means a lower growth rate of emissions as these economies develop. As incomes rise in developing countries, their populations will increasingly demand transportation services that capture the immense societal benefits transportation brings. The majority of growth in transportation emissions over the past several decades has occurred in the light-duty sector in developing countries, and this will likely be the main source of future growth. Developing countries' emissions growth can swamp reductions in developed countries. As an illustration, a 50 percent reduction in on-road transportation emissions by developed countries relative to current emissions would be *completely offset* by just eight years of growth in on-road transportation emissions in the developing world (assuming the continuation of 4.4 percent per annum growth rate experienced since 1970).

Here, two areas of innovation are important. Conditional on growing demand for transportation, it will be necessary to reduce the carbon intensity. Although, to date, attention in this area has focused on solutions that leave the fundamental concept of personal transportation unchanged (as in the electrification of traditional passenger vehicles), innovation in the developing world might move in novel directions. One such example are the electric rickshaws with swappable batteries that have grown quickly in India and allow for electrification while avoiding the challenges of household-level charging (Schmall and Ewing 2022). Second, some quickly growing cities in the developing world may offer opportunities for novel approaches to urban planning, to direct urban development towards reducing transportation needs or strategically siting high-density development along public transport corridors (Nakamura and Hayashi 2013). Admittedly, this problem is likely to be a challenging one to solve, given the strong historical links between transportation demand and income. In the developed world, such opportunities are already constrained by existing (vehicle-based) infrastructure (Glaeser and Kahn 2010). High transportation demand growth in the developing world thus presents not just a challenge, but an opportunity.

Maintaining Flexibility

Existing technology is not sufficient. Deep decarbonization of the transportation sector depends on continued innovation and technological progress, and predicting the direction and pace of innovation is somewhere between difficult and impossible. It follows that the optimal policy pathway maintains the viability of as many technology pathways as possible, and avoids prematurely "picking winners."

History is replete with examples of both overly pessimistic and overly optimistic assessments of environmental innovation. For instance, forecasted compliance costs of the Acid Rain Program exceeded estimates made after the policy took effect by a factor of five (Chan et al. 2012) by dramatically underestimating the ability of industry to adjust in response. On the other hand, despite substantial government subsidies and a federal mandate in the Energy Independence and Security Act of 2007 that advanced biofuels like cellulosic ethanol would constitute roughly half of biofuel production by 2022, progress in this area has been elusive and cellulosic ethanol remains uncompetitive on a cost-basis with other fuels (Chen et al. 2021).

Technological progress sometimes proceeds smoothly with a series of incremental gains to an established technology; at other times, innovation can be lumpy and discontinuous. As one example, hydraulic fracturing led to a doubling of US natural gas production in the past 15 years, whereas just before that time, the Annual Energy Outlook predicted stable domestic natural gas production and increasing US reliance on imports (EIA 2008). This resulting rapid expansion of natural gas production has facilitated some decarbonization of the US electric grid, improving the emissions profiles of electric vehicles (Holland et al. 2020). The unpredictable and lumpy nature of technological progress highlights the benefits of technology-neutral policies that do not foreclose potential decarbonization pathways. Although electrification is, based on current technology, the most obvious pathway to reduce emissions from light-duty vehicles, innovation may offer lower-cost pathways in the future.

In a similar vein, decarbonization of transportation in developing world will rely on continued innovation, along potentially novel directions. Solutions, such as widespread vehicle electrification, may work well for some sectors or regions, but may not be able to address unique challenges in other settings. Innovation along other pathways (like biofuels) might ultimately provide the best prospects for, say, decarbonizing on-road transportation in developing countries.

One challenge is that the majority of energy innovation occurs in a handful of countries—the United States, Japan, China, Korea, and countries of the European Union—and roughly three-quarters of energy research and development spending is incurred by the private sector (IEA 2020). A long literature in environmental economics documents how policy can induce innovation along specific pathways (for example, Newell, Jaffe, and Stavins 1999; Grubb et al. 2021). If policy in developed countries focuses innovation along domestic pathways (like vehicle electrification), decarbonization in emerging economies might occur more slowly. Similar concerns have long been recognized for pharmaceutical innovation, where market and policy combine to slow innovation for therapeutics for less-affluent patients (Pécoul et al. 1999). Widespread decarbonization will almost surely involve continued progress across a wide range of potential pathways. Technologically-neutral policies are thus preferable. A carbon price combined with subsidies for primary research may ultimately produce new technologies that are beneficial both for specific uses in developed countries and the future needs of developing countries.

Solving Problems of Collective Action

Decarbonization is unlikely to succeed without addressing the collective action problems inherent in carbon markets. One (obvious) challenge to collective action is that the environmental costs and benefits associated with climate change, spillovers from research, and economies of scale in production all extend beyond the political and economic boundaries of nations (Das Gupta 2014). The political economy of decarbonization has long posed challenges within and across countries. It is fraught with ethical arguments about the responsibilities of countries that developed through the use of carbon emissions and often pits winners and losers from abatement policy against each other.

Yet political motivation seems higher now than in the past. At the country level, policymakers have enacted policies to speed the energy transition: for examples from the United States, the Inflation Reduction Act of 2022 and the Infrastructure Investment and Jobs Act of 2021 both subsidize decarbonization efforts in different ways. Internationally, a growing number of nations have joined the Net-Zero Coalition, with countries that currently account for roughly three-quarters of global emissions pledging to reach carbon-neutrality. Business, educational institutions, and other organizations have joined the UN Race to Zero, with the goal of halving carbon emissions by 2030.

Despite the apparent progress, we note two sources of context for the momentum of the past few years. First, although many countries have pledged to reduce their carbon emissions, the 2018 report from the Intergovernmental Panel on Climate Change assesses that the aggregate pledges to date are either too small or insufficiently prompt to limit temperature increases to 1.5°C by the end of the twenty-first century. The Intergovernmental Panel on Climate Change notes emissions must fall by 45 percent by 2030, while current commitment plans allow for 10 percent growth in emissions over the period (United Nations Climate Change 2022).

Second, financial support for developing countries has generally been insufficient relative to the anticipated costs. UN estimates adaptation costs for developing countries to exceed $300 billion per year by 2030 (United Nations 2021), the cost of the United Nations Sustainable Development Goals at $5–7 trillion over 2015–2030, and the gap in infrastructure funding worldwide at $15 trillion through 2040 (Economics 2017). Financial commitments from developed countries were the focus of the United Nations Climate Change Conference (COP27) held in November 2022 and strike at the heart of ethical arguments about the responsibility of the developed world to compensate developing countries for climate damages and subsidize decarbonization in lower-income countries. Although developed countries have increasingly made monetary commitments to assist developing countries, the aggregate commitments have fallen short on a $100 billion per annum climate finance target, despite a high fraction of the funding being offered as loans rather than grants to developing countries (Timperley 2021).

A technological solution to the collective action challenge is also being developed: direct air capture. These technologies extract carbon dioxide directly from the atmosphere. Unilateral deployment would yield benefits for the entire planet in the same way as the global inventory of emissions determines the level and rate of warming. If direct air capture were to become economically viable at scale, it would introduce the prospect of a successful climate change mitigation path that supports a higher level of long-run emissions. That is, some level of decarbonization defection could be supported. Current pilot projects are small and expensive, but the potential benefits are sufficiently high that these efforts should be expanded.

Mitigating the Political Costs of Action

The public and political appetites for bold climate action are implicitly predicated on continued access to inexpensive energy and transportation services. The substantial increase in energy prices in general during 2022, and transportation fuels in particular, increased pressure on governments around the world to lower prices and increase supply—even at the expense of substantially increasing carbon emissions. For example, high US gasoline prices led a number of states to enact temporary moratoria for state gas taxes and for roughly one-third of the oil to be withdrawn from the Strategic Petroleum Reserve. European countries enacted electricity price caps. Concerns about the reliability of natural gas supplies led Germany and other European countries to restart previously decommissioned coal-fired power plants in the past few months (Morris 2022).

Actions by developed countries have cascaded down to developing countries. Voracious European demand for liquified natural gas as a substitute for Russian natural gas pushed many developing countries towards older, higher carbon sources of energy (Tani and Parkin 2022). Sanctions on Russia have been repeatedly diluted to maintain the flow of Russian oil and refined products into the world market. India and China have snapped up imports of these discounted Russian products over the past twelve months (Menon 2023).

Although future cost reductions in green technologies might soften the economic blow of climate-friendly policies, revealed preference suggests that climate concerns take a back seat to lower energy prices for citizens and policymakers alike. As we have seen time and time again, it is the politics of carbon abatement, not the policy of carbon abatement, that has most stymied progress towards a cleaner global transportation sector.

Hard Truths

Policymakers wishing to decarbonize the transportation sector face a menu of unappealing options. We remain in a phase of technology development characterized by significant uncertainty about the optimal path in all sectors. Governments worldwide, to the extent they are taking action at all, have overwhelmingly chosen the path of "carrots," not "sticks." In the absence of several favorable draws from innovation lotteries, this pathway will likely be expensive and characterized by only partial decarbonization success. Electrification, today's preferred technology, faces numerous obstacles, each of which will have to be overcome to make this pathway environmentally transformative while remaining affordable. Developing economies are particularly ill-suited to electrification. One task in society is to figure out where and how hard to push forward with electrification. Fortunately, economists know how to set incentives that will help guide resource allocation in this environment.

Raising the price of pollution remains the most important, and likely necessary, approach to decoupling growth from emissions (the merits of which are surveyed by Knittel 2012). Unfortunately, it is out of favor in many places. Governments have instead turned to subsidizing "green" alternatives. Even if the green alternatives were carbon-free (which they typically are not), subsidies for green technology are not equivalent to taxes on pollution. In at least one important sense, the subsidy approach is counterproductive. Subsidy-favored technologies become artificially inexpensive to adopt, which expands overall demand while crowding out profitable innovation along currently unfavored or not-yet-imagined abatement pathways. The opportunity cost is incalculable. The countries of Africa offer a concrete example of this concern. Their population will likely double in the next century, and transportation demand will increase in concert with a larger and richer population. It will be advantageous for urban planning to center around public transit and small vehicles in these economies. Increasing the cost of pollution creates incentives for cleaner urban growth, but cheap electric vehicles does not.

With these broader issues and options in mind, policymakers should recommit to implementing policies that set the right incentives. (Economists have an essential advisory role to play here.) Climate policymakers would be well-served by extending their time horizon to reflect the fact that decarbonizing transportation will be a multi-decade project. Framing decarbonization as necessary to occur by <insert your preferred net-zero date here> undermines credibility if we continue to miss "point-of-no-return" deadlines. It also risks locking us into the set of presently-feasible technology options. Similarly, all-or-nothing targets ("100 percent <insert preferred technology here>") and thresholds ("1.5 degrees . . .") may impose high costs of abatement, or achieve lower than expected levels of abatement, by failing to equate social costs and benefits on the margin. This is especially true with respect to abating the last units of pollution, or with converting the last users to green technology if green and brown technologies are imperfect substitutes (Holland et al. 2020). In short, a return to basic economic principles is desirable.

The world is on a cusp of a transformational shift in how we move goods and people. We must find a way to do so while continuing to respect the immense value of the underlying transportation services. A defining challenge will be to develop and select technologies that reduce carbon emissions from transportation sectors that have starkly different needs. How to decarbonize light-duty vehicles is particularly important. From a global perspective, successful carbon-reducing innovation must push against the increase in emissions from transportation demand that results from developing countries entering the middle class. We are in early days of a long transition, and humility about which technology pathways will ultimately satisfy the needs of each sector is appropriate.

■ *We thank Reid Taylor and Jessica Lyu for excellent research assistance. All opinions and errors are our own. The views expressed here are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Dallas or the Federal Reserve System.*

References

Allain, Marc, David Atherton, Igor Gruden, Sandeep Singh, and Kevin Sisken. 2012. "Daimler's Super Truck Program; 50% Brake Thermal Efficiency." Unpublished.

Allcott, Hunt, Allan Collard-Wexler, and Stephen D. O'Connell. 2016. "How Do Electricity Shortages Affect Industry? Evidence from India." *American Economic Review* 106 (3): 587–624.

Allcott, Hunt, and Michael Greenstone. 2012. "Is There an Energy Efficiency Gap?" *Journal of Economic Perspectives* 26 (1): 3–28.

Allcott, Hunt, and Michael Greenstone. 2017. "Measuring the Welfare Effects of Residential Energy Efficiency Programs." NBER Working Paper 23386.

Archsmith, James, Erich Muehlegger, and David S. Rapson. 2021. "Future Paths of Electric Vehicle Adoption in the United States: Predictable Determinants, Obstacles and Opportunities." *Environmental and Energy Policy and the Economy* 3: 71–110.

Blimpo, Moussa, Shaun McRae, and Jevgenijs Steinbuks. 2018. "Why Are Connection Charges So High? An Analysis of the Electricity Sector in Sub-Saharan Africa." World Bank Policy Research Working Paper 8407.

Cai, Ximing, Xiao Zhang, and Dingbao Wang. 2011. "Land Availability for Biofuel Production." *Environmental Science and Technology* 45 (1): 334–39.

Chan, Gabriel, Robert Stavins, Robert Stowe, and Richard Sweeney. 2012. "The SO2 Allowance-Trading System and the Clean Air Act Amendments of 1990: Reflections on 20 Years of Policy Innovation." *National Tax Journal* 65 (2): 419–52.

Chen, Luoye, Deepayan Debnath, Jia Zhong, Kelsie Ferin, Andy VanLoocke, and Madhu Khanna. 2021. "The Economic and Environmental Costs and Benefits of the Renewable Fuel Standard." *Environmental Research Letters* 16 (3): 034021.

Condon, Nicole, Heather Klemick, and Ann Wolverton. 2015. "Impacts of Ethanol Policy on Corn Prices: A Review and Meta-analysis of Recent Evidence." *Food Policy* 51: 63–73.

Congressional Research Service. 2022. "Sustainable Aviation Fuel (SAF): In Brief." CRS Report R47171.

Covert, Thomas, Michael Greenstone, and Christopher R. Knittel. 2016. "Will We Ever Stop Using Fossil Fuels?" *Journal of Economic Perspectives* 30 (1): 117–38.

Das Gupta, Monica. 2014. "Population, Poverty, and Climate Change." *World Bank Research Observer* 29 (1): 83–108.

Dasgupta, Susmita, Benoit Laplante, Hua Wang, and David Wheeler. 2002. "Confronting the Environmental Kuznets Curve." *Journal of Economic Perspectives* 16 (1): 147–68.

Davis, Lucas W. 2019. "Evidence of a Homeowner-Renter Gap for Electric Vehicles." *Applied Economics Letters* 26 (11): 927–32.

Davis, Lucas W., Catherine Hausman, and Nancy L. Rose. 2023. "Transmission Impossible? Prospects for Decarbonizing the US Grid." *Journal of Economic Perspectives* 37 (4): forthcoming.

Davis, Stacy C., and Robert G. Boundy. 2022. "Transportation Energy Data Book." Oak Ridge National Laboratory. https://tedb.ornl.gov/data/2022 (accessed March 30, 2022).

Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." *American Economic Review* 101 (7): 3078–108.

Elmallah, Salma, Anna M. Brockway, and Duncan Callaway. 2022. "Can Distribution Grid Infrastructure Accommodate Residential Electrification and Electric Vehicle Adoption in Northern California?" *Environmental Research: Infrastructure and Sustainability* 2 (4): 045005.

Energy Information Administration (EIA). 2008. *Annual Energy Outlook 2008*. Washington, DC: U.S. Energy Information Administration.

Energy Information Administration (EIA). 2023. *Annual Energy Outlook 2023*. Washington, DC: U.S. Energy Information Administration.

Esau, Steve. 2021. "Fuel Comparisons Must Consider Energy Density." *The Maritime Executive*, September 17. https://maritime-executive.com/editorials/future-fuel-comparisons-must-consider-energy-density.

European Commission. 2023. "EDGAR-Emissions Database for Global Atmospheric Research. Global Greenhouse Gas Emissions." European Commission. https://edgar.jrc.ec.europa.eu/dataset_ghg50 (accessed August 2, 2022).

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram. 2018. "Do Energy Efficiency Investments Deliver? Evidence from the Weatherization Assistance Program." *Quarterly Journal of Economics* 133 (3): 1597–1644.

Gillingham, Kenneth, David Rapson, and Gernot Wagner. 2016. "The Rebound Effect and Energy Efficiency Policy." *Review of Environmental Economics and Policy* 10 (1): 68–88.

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

Grossman, Gene M., and Alan B. Krueger. 1991. "Environmental Impacts of a North American Free Trade Agreement." NBER Working Paper 3914.

Grubb, Michael, Paul Drummond, Alexandra Poncia, Will McDowall, David Popp, Sascha Samadi, Cristina Penasco et al. 2021. "Induced Innovation in Energy Technologies and Systems: A Review of Evidence and Potential Implications for CO2 Mitigation." *Environmental Research Letters* 16 (4): 043007.

Havlík, Petr, Uwe A. Schneider, Erwin Schmid, Hannes Böttcher, Steffen Fritz, Rastislav Skalský, Kentaro Aoki et al. 2011. "Global Land-Use Implications of First and Second Generation Biofuel Targets." *Energy Policy* 39 (10): 5690–5702.

Hill, Jason, Erik Nelson, David Tilman, Stephen Polasky, and Douglas Tiffany. 2006. "Environmental, Economic, and Energetic Costs and Benefits of Biodiesel and Ethanol Biofuels." *Proceedings of the National Academy of Sciences* 103 (30): 11206–10.

Holladay, Johnathan, Zia Abdullah, and Joshua Heynu. 2020. *Sustainable Aviation Fuel: Review of Technical Pathways*. Washington, DC: U.S. Department of Energy Office of Energy Efficiency and Renewable Energy.

Holland, Stephen P., Jonathan E. Hughes, and Christopher R. Knittel. 2009. "Greenhouse Gas Reductions under Low Carbon Fuel Standards?" *American Economic Journal: Economic Policy* 1 (1): 106–46.

Holland, Stephen P., Erin T. Mansur, Nicholas Z. Muller, and Andrew J. Yates. 2020. "Decompositions and Policy Consequences of an Extraordinary Decline in Air Pollution from Electricity Generation." *American Economic Journal: Economic Policy* 12 (4): 244–74.

Intergovernmental Panel on Climate Change. 2014. "Transport." In *Climate Change 2014: Mitigation of Climate Change*, edited by Ottmar Edenhofer et al., 599–670. Cambridge, UK: Cambridge University Press.

International Energy Agency (IEA). 2019. *How Competitive Is Biofuel Production in Brazil and the United States? Analysis from Renewables 2018*. Paris: International Energy Agency.

International Energy Agency (IEA). 2020. *Global Status of Clean Energy Innovation in 2020*. Paris: International Energy Agency.

International Energy Agency (IEA). 2021. *Ammonia Technology Roadmap*. Paris: International Energy Agency.

International Energy Agency (IEA). 2022a. "Energy Statistics Data Browser." IEA. https://www.iea.org/data-and-statistics/ data-tools/energy-statistics-data-browser 2022 (accessed March 21, 2022).

International Energy Agency (IEA). 2022b. *Global Energy Review: CO2 Emissions in 2021.* Paris: International Energy Agency.

International Energy Agency (IEA). 2022c. *Global Supply Chains of EV Batteries.* Paris: International Energy Agency.

International Energy Agency (IEA). 2022d. *How Competitive is Biofuel Production in Brazil and the United States? Analysis from Renewables 2018.* Paris: International Energy Agency.

International Energy Agency (IEA). 2022e. *International Shipping.* Paris: International Energy Agency.

International Energy Agency (IEA). 2022f. *Renewables 2022.* Paris: International Energy Agency.

International Energy Agency (IEA). 2023. *Global EV Outlook 2023.* Paris: International Energy Agency.

International Maritime Organization, "2021 Guidelines on the Shaft/Engine Power Limitation System to Comply with the EEXI Requirements and Use of a Power Reserve," MEPC.335(76), Marine Environmental Protection Committee, 17 June 2021.

Jacobsen, Mark R. 2013. "Evaluating US Fuel Economy Standards in a Model with Producer and Household Heterogeneity." *American Economic Journal: Economic Policy* 5 (2): 148–87.

Jacobsen, Mark, and Arthur A. van Benthem. 2015. "Vehicle Scrappage and Gasoline Policy." *American Economic Review* 105 (3): 1312–38.

Kara, Siddharth. 2023. *Cobalt Red: How the Blood of the Congo Powers Our Lives.* New York: St. Martin's Press.

Kargul, John, Mark Stuhldreher, Daniel Barba, Charles Schenk, Stanislav Bohac, Joseph McDonald, and Paul Dekraker. 2019. "Benchmarking a 2018 Toyota Camry 2.5-Liter Atkinson Cycle Engine with Cooled-EGR." *SAE International Journal of Advances and Current Practices in Mobility* 1 (2): 601–38.

Keeney, Roman, and Thomas W. Hertel. 2009. "The Indirect Land Use Impacts of United States Biofuel Policies: The Importance of Acreage, Yield, and Bilateral Trade Responses." *American Journal of Agricultural Economics* 91 (4): 895–909.

Knittel, Christopher R. 2011. "Automobiles on Steroids: Product Attribute Trade-Offs and Technological Progress in the Automobile Sector." *American Economic Review* 101 (7): 3368–99.

Knittel, Christopher R. 2012. "Reducing Petroleum Consumption from Transportation." *Journal of Economic Perspectives* 26 (1): 93–118.

Lee, Jordy, Morgan Bazilian, B. Sovacool, and S. Greene. 2020. "Responsible or Reckless? A Critical Review of the Environmental and Climate Assessments of Mineral Supply Chains." *Environmental Research Letters* 15 (10): 103009.

Lee, Kenneth, Edward Miguel, and Catherine Wolfram. 2016. "Appliance Ownership and Aspirations among Electric Grid and Home Solar Households in Rural Kenya." *American Economic Review* 106 (5): 89–94.

Melillo, Jerry M., John M. Reilly, David W. Kicklighter, Angelo C. Gurgel, Timothy W. Cronin, Sergey Paltsev, Benjamin S. Felzer, Xiaodong Wang, Andrei P. Sokolov, and C. Adam Schlosser. 2009. "Indirect Emissions from Biofuels: How Important?" *Science* 326 (5958): 1397–99.

Menon, Shruti. 2023. "Ukraine Crisis: Who Is Buying Russian Oil and Gas." *BBC*, May 23. https://www.bbc.com/news/world-asia-india-60783874 (accessed December 6, 2022).

Morris, Loveday. 2022. "Germany Is Firing Up Old Coal Plants, Sparking Fears Climate Goals Will Go Up in Smoke." *Washington Post*, August 1. https://www.washingtonpost.com/world/2022/08/01/germany-coal-energy-climate/.

Nakamura, Kazuki, and Yoshitsugu Hayashi. 2013. "Strategies and Instruments for Low-Carbon Urban Transport: An International Review on Trends and Effects." *Transport Policy* 29: 264–74.

National Academies of Sciences, Engineering, and Medicine. 2016. *Commercial Aircraft Propulsion and Energy Systems Research: Reducing Global Carbon Emissions.* Washington, DC: National Academies Press.

National Academies of Sciences, Engineering, and Medicine. 2021. *Assessment of Technologies for Improving Light-Duty Vehicle Fuel Economy 2025-2035.* Washington, DC: National Academies Press.

Newell, Richard G., Adam B. Jaffe, and Robert N. Stavins. 1999. "The Induced Innovation Hypothesis and Energy-Saving Technological Change." *Quarterly Journal of Economics* 114 (3): 941–75.

Nikolaidis, Pavlos, and Andreas Poullikkas. 2017. "A Comparative Overview of Hydrogen Production Processes." *Renewable and Sustainable Energy Reviews* 67: 597–611.

Nykvist, Björn, and Måns Nilsson. 2015. "Rapidly Falling Costs of Battery Packs for Electric Vehicles." *Nature Climate Change* 5 (4): 329–32.

Oxford Economics. 2017. *Global Infrastructure Outlook: Forecasting Infrastructure Investment Needs and Gaps.* Sydney: Global Infrastructure Hub.

Pascale, Andrew, Jesse D. Jenkins, and Emily Leslie. 2021. *Princeton's Net-Zero America Study: Annex F: Integrated Transmission Line Mapping and Costing.* Princeton, NJ: Princeton University Press.

Pécoul, Bernard, Pierre Chirac, Patrice Trouiller, and Jacques Pinel. 1999. "Access to Essential Drugs in Poor Countries: A Lost Battle?" *Journal of the American Medical Association* 281 (4): 361–67.

Qiao, Qinyu, Fuquan Zhao, Zongwei Liu, Xin He, and Han Hao. 2019. "Life Cycle Greenhouse Gas Emissions of Electric Vehicles in China: Combining the Vehicle Cycle and Fuel Cycle." *Energy* 177: 222–33.

Rapson, David and Erich Muehlegger. 2023. "Replication data for: Global Transportation Decarbonization." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E191761V1.

Rapson, David, and James Bushnell. Forthcoming. "The Electric Ceiling: Limits and Costs of Full Electrification." *Review of Environmental Economics and Policy***.**

Roberts, Michael J., and Wolfram Schlenker. 2013. "Identifying Supply and Demand Elasticities of Agricultural Commodities: Implications for the US Ethanol Mandate." *American Economic Review* 103 (6): 2265–95.

Rodrigue, Jean-Paul. 2020. *The Geography of Transport Systems*. 5th ed. Abingdon, UK: Routledge.

Ryan, Nicholas. 2020. "Contract Enforcement and Productive Efficiency: Evidence from the Bidding and Renegotiation of Power Contracts in India." *Econometrica* 88 (2): 383–424.

Schlachter, Fred. 2012. "Has the Battery Bubble Burst?" *APS News* 21 (8): 8.

Schmall, Emily, and Jack Ewing. 2022. "India's Electric Vehicle Push Is Riding on Mopeds and Rickshaws." *New York Times*, September 4. https://www.nytimes.com/2022/09/04/business/energy-environment/india-electric-vehicles-moped-rickshaw.html.

Searchinger, Timothy, Ralph Heimlich, Richard A. Houghton, Fengxia Dong, Amani Elobeid, Jacinto Fabiosa, Simla Tokgoz, Dermot Hayes, and Tun-Hsiang Yu. 2008. "Use of U.S. Croplands for Biofuels Increases Greenhouse Gases through Emissions from Land-Use Change." *Science* 319 (5867): 1238–40.

Shapiro, Joseph S. 2016. "Trade Costs, CO2, and the Environment." *American Economic Journal: Economic Policy* 8 (4): 220–54.

Statista. 2020. "Energy Density of Maritime Fuels in 2020, by Fuel Type (in Megajoule per Liter)." https://www.statista.com/statistics/1279431/energy-density-of-maritime-fuels/ (accessed December 2, 2022).

Tani, Shotaro, and Benjamin Parkin. 2022. "Europe's Appetite for LNG Leaves Developing Nations Starved of Gas." *Financial Times*, September 23. https://www.ft.com/content/752b1285-3174-4cf1-83c0-b1151888bf4e.

Timperley, Jocelyn. 2021. "The Broken $100-Billion Promise of Climate Finance—and How to Fix It." *Nature* 598 (7881): 400–402.

Tran, Dat T., Jessica M. Palomino, and Scott R. J. Oliver. 2018. "Desulfurization of JP-8 Jet Fuel: Challenges and Adsorptive Materials." *RSC Advances* 8 (13): 7301–14.

United Nations. 2021. "Guterres Urges Developed Countries to Deliver on Climate Pledge for Vulnerable Nations." *UN News*, November 2. https://news.un.org/en/story/2021/11/1104712.

United Nations Climate Change. 2022. "Climate Plans Remain Insufficient: More Ambitious Action Needed Now." *UN Climate Change News*, October 26. https://unfccc.int/news/climate-plans-remain-insufficient-more-ambitious-action-needed-now.

United Nations Department of Economic and Social Affairs. 2022. *World Population Prospects 2022: Summary of Results*. New York: United Nations.

US Department of Energy. 2021. "Alternative Fuels Data Center Fuel Properties Comparison." US Department of Energy. https://afdc.energy.gov/fuels/properties (accessed December 2, 2022).

US Department of Energy. 2023a. "Alternative Fuels Data Center Alternative Fueling Station Locator." US Department of Energy. https://afdc.energy.gov/stations/#/find/nearest (accessed November 3, 2022).

US Department of Energy. 2023b. "Hydrogen Storage." US Department of Energy.

US Department of Transportation. 2019. "What the Cost of Airline Fuel Means to You." September 13. https://www.transportation.gov/administrations/assistant-secretary-research-and-technology/what-cost-airline-fuel-means-you.

US Environmental Protection Agency (EPA). 2023. "Lifecycle Greenhouse Gas Results." US EPA. https://www.epa.gov/fuels-registration-reporting-and-compliance-help/lifecycle-greenhouse-gas-results.

Wakabayashi, Daisuke, and Claire Fu. 2022. "For China's Auto Market, Electric Isn't the Future. It's the Present." *New York Times*, September 26. https://www.nytimes.com/2022/09/26/business/china-electric-vehicles.html.

Witcover, Julie, and Robert B. Williams. 2020. "Comparison of 'Advanced' Biofuel Cost Estimates: Trends during Rollout of Low Carbon Fuel Policies." *Transportation Research Part D: Transport and Environment* 79: 102211.

World Bank. 2019. "World Economic Forum Global Competitiveness Index." World Bank. https://tcdata360.worldbank.org (accessed April 1, 2022).

World Bank. 2022. *Global Economic Prospects 2022*. Washington, DC: World Bank.

World Bank. 2023. "GDP per Capita (Current US$)." World Bank. https://data.worldbank.org/indicator/NY.GDP.PCAP.CD (accessed April 1, 2019).

Journal of Economic Perspectives—Volume 37, Number 3—Summer 2023—Pages 189–212

Compensating Wage Differentials in Labor Markets: Empirical Challenges and Applications

Kurt Lavetti

How do workers choose which jobs to accept? Although earnings are one important factor in this decision, workers typically also consider many other job characteristics. They may be willing to accept jobs that pay less but offer flexible hours, health insurance, or shorter commutes. Dangerous or dirty jobs like coal mining or trash collection may have to offer higher pay to entice workers to accept the undesirable characteristics of the job. Of course, people may disagree about whether a particular job characteristic is good or bad. Some people prefer working at a desk, while others prefer being outside or doing physically active work.

Conceptually, choosing a job can be thought of as a worker selling their services in the labor market and simultaneously buying amenities (which can be positive or negative) from their employer. The observed wage rate combines the value of a worker's time and the implicit prices of all amenities. Quantifying the tradeoffs between earnings and job characteristics, also called "compensating wage differentials," is of fundamental importance for understanding labor market equilibria and wage dispersion.

To illustrate, consider the differences in average wages and education levels across occupations shown in Table 1. Notice that although bakers have six times higher college graduation rates than butchers, they earn about 8 percent less on average. This may be, in part, because most people would prefer, all else equal, to work in a redolent bakery full of fresh bread than in a butchery, so butchers must be paid extra for the undesirable characteristics of their job. And although accountants

■ *Kurt Lavetti is Associate Professor of Economics, The Ohio State University, Columbus, Ohio and a Research Associate at the National Bureau of Economic Research. His email address is lavetti.1@osu.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.3.189.

Table 1
Average Education and Annual Full-Time Earnings of Selected Occupations

Occupation	*High school degree*	*College degree*	*Annual earnings*
Butcher	0.70	0.03	$36,150
Baker	0.84	0.18	$33,570
Accountant	1.00	0.79	$72,020
Explosives handler	0.92	0.00	$100,120
Forest conservationist	0.97	0.39	$36,760
Agricultural inspector	0.94	0.39	$49,460

Source: IPUMS March CPS data (Flood et al. 2022).
Note: Averages are calculated using March CPS data from 2012–2022. Columns 1 and 2 report the average share of workers by occupation with at least a high school degree and at least a college degree, respectively. Column 3 reports average annual earnings of a full-time equivalent worker by occupation using 2020 dollars.

are highly educated, they earn 28 percent less than explosives handlers, who have low college completion rates but are exposed to physical danger. Finally, forest conservationists earn 26 percent less than agricultural inspectors despite similar education levels. Workers who choose forest conservation jobs may be willing to accept lower salaries in exchange for the opportunity to work outdoors, frequently in scenic parks, and also to feel good about contributing to environmental stewardship. These are just examples, but any feature of a job that workers have preferences over could potentially have an associated compensating wage differential.

Although the concept of a compensating differential is straightforward and intuitive, its simplicity belies how difficult it is to empirically quantify compensating differentials in real-world labor markets. Although Adam Smith (1776) wrote about the tradeoffs between earnings and job characteristics,[1] it would take nearly 200 years until Sherwin Rosen (1974; 1986) formalized a theoretical and empirical framework for studying compensating differentials. In this article, I begin with an overview of the Rosen (1974; 1986) model, which emphasizes that compensating wage differentials result from workers with different preferences for amenities sorting between firms with different costs of providing amenities.

I then present a chronology of the empirical approaches used to estimate compensating differentials, highlighting the new lessons learned as data quality and methods advanced over time and how these advances in turn revealed new challenges and setbacks. I begin with basic cross-sectional wage models, and then discuss how panel data models were used to alleviate bias caused by unobserved worker skills. Attempts to extend the theory of compensating wage differentials to markets with imperfect competition or search frictions led to grave concerns about

[1] In *The Wealth of Nations* (Book X, Part I), Adam Smith (1776) writes "[T]he wages of labour vary with the ease or hardship, the cleanliness or dirtiness, the honourableness or dishonourableness of the employment. A . . . blacksmith . . . seldom earns so much . . . as a [coal miner] . . . . His work is not quite so dirty, is less dangerous, and is carried on in daylight, and above ground. . . ."

the applicability of the Rosen model to realistic labor market settings and sowed doubt about the reliability of empirical estimates. Recent progress in responding to these concerns has involved using newly available matched worker-firm data and settings with quasi-random variation in levels of job amenities.

Empirical estimates of compensating differentials are important for informing and designing a broad range of public policies. One example is that estimates of the compensating differential for the risk of death on a job are used to calculate the value of statistical life, which in turn is widely used in cost-benefit analyses by government regulators. Compensating differentials are also important for interpreting and measuring earnings disparities, which inform public policies aimed at reducing inequality. For example, Sorkin (2018) estimates that compensating wage differentials explain at least 15 percent of earnings inequality in the United States, while Taber and Vejlin (2020) estimate that they explain up to 26 percent of wage inequality in Denmark. To put the importance of compensating differentials in a more direct perspective, the authors estimate that if all amenities were eliminated from all jobs, total wages would increase by about 18 percent. In other words, compensating differentials are roughly as important for aggregate economic activity as the entire healthcare sector is for the US economy.

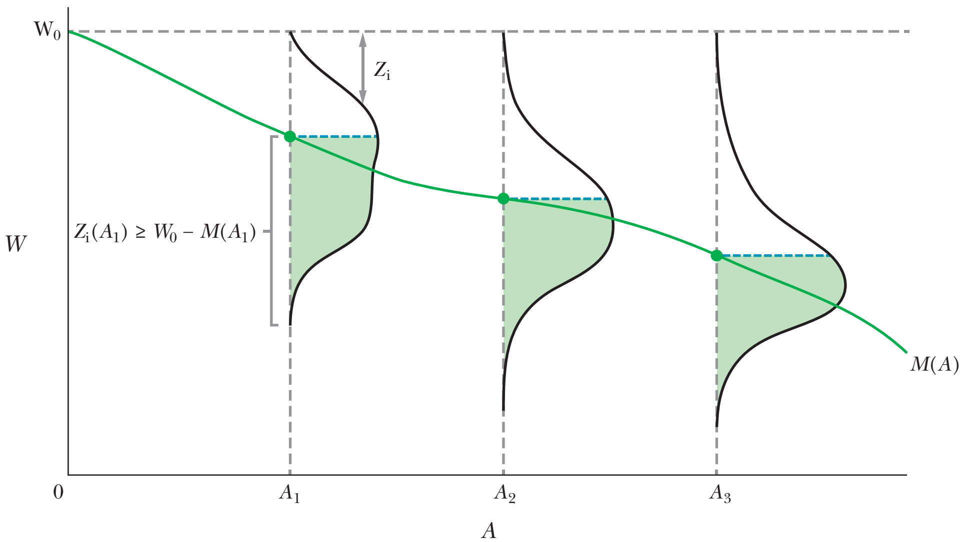
The Rosen Model: Parameters and Interpretations

The Rosen (1986) model provides a theoretical framework for understanding tradeoffs between wages and amenities in labor markets. Consider a competitive labor market in which workers choose between jobs that offer different levels of wages, W, and amenities, A, to maximize utility $u_i(W, A)$. A includes both good and bad job characteristics, where a reduction in disamenities can be expressed as an increase in amenities, and vice versa. The subscript on u_i indicates that different workers, denoted by i, have different preferences for amenities relative to wages.

In this model, all workers are assumed to have the same productivity—the model abstracts from factors other than A that might cause wages to vary, making it important to control these other factors in empirical estimation. Workers choose whether to sell their services in the labor market at wage rate W and also choose how much A to buy, where the price of A is an implicit reduction in the wage rate. In this sense, a worker's decision to accept a job can be viewed as a simultaneous selling and buying choice. The observed wage is the sum of these two transactions that are tied together but conceptually distinct.

Let's begin by focusing on the worker's side of the labor market in a basic discrete version of the model, in which jobs offer either zero amenities or fixed levels A_1, A_2, or A_3. Examples of discrete amenities include the number of days per week employees can work from home or weeks of vacation time per year. Workers who choose to accept a job with zero amenities earn the market wage, denoted W_0. Jobs with more amenities can offer lower wages and still attract workers, creating a downward-sloping relationship between equilibrium wages and amenities, as shown in the

Figure 1.
Willingness to Pay for Amenities and Job Choices



Source: Figure created by author.
Note: The horizontal axis shows levels of amenities A. The vertical axis shows wages W. W_0 is the wage for a job with zero amenities. For each discrete level of amenities A_1, A_2, and A_3, the Z_i curve shows the distribution of preferences for the given amenity level across individuals. The $M(A)$ function shows the downward-sloping relationship between wages and amenities in equilibrium. All the workers in the shaded region, for whom $Z_i(A) > W_0 - M(A)$, value amenities by more than the market reduction in wages required to obtain them.

$M(A)$ function in Figure 1. The vertical distance between W_0 and the $M(A)$ function is how much of their potential wage workers give up in exchange for amenities.

Although there is a single market wage for jobs that provide A_1, workers have differing preferences. Imagine that each worker was asked: what is the maximum reduction in wages, Z_i, you would accept to obtain a job with amenity level A_1? Because workers have heterogeneous preferences, there is a distribution of Z_is. The solid black line in Figure 1 shows this distribution (oriented along the vertical axis). Workers who value amenities the most have high Z_is, indicating a high willingness to give up wages in exchange for amenities. These workers are located on the bottom tail of the distribution, furthest from W_0. Those who do not value amenities at all have $Z_i = 0$ and are located at the top end of the black line, where $W = W_0$. As the figure suggests, the distributions of Z_is could be arbitrary, but for simplicity consider the case where $Z_i \geq 0$, as the equilibrium is more complicated if workers disagree about whether A is good or bad. The $M(A)$ function splits workers into two groups: those with small Z_is are unwilling to accept the market wage $M(A)$ and prefer to take the job with zero amenities, while those with large Z_is (the shaded region) prefer to take a job with amenities.

A key feature of this model is that workers who hold jobs with amenities earn (weakly) positive surplus, or rents, from these jobs. All the workers in the shaded region, for whom $Z_i(A) > W_0 - M(A)$, value amenities by more than the market reduction in wages required to obtain them. Because of rents, if two workers who chose jobs with different levels of A were forced to switch jobs with each other, total surplus would generally fall. This is unlike equilibria in most economic models of competitive markets, where the matching between particular buyers and sellers is not relevant to total welfare. The welfare effects of rents in the Rosen model are similar to the effects of bundling jobs with purchases of goods that yield positive consumer surplus.

The presence of rents also has implications that extend beyond the Rosen model. For example, in models of equilibrium wages that are based on negotiations between workers and firms, rents can impact the outcomes of negotiations, affecting wages and job mobility patterns.

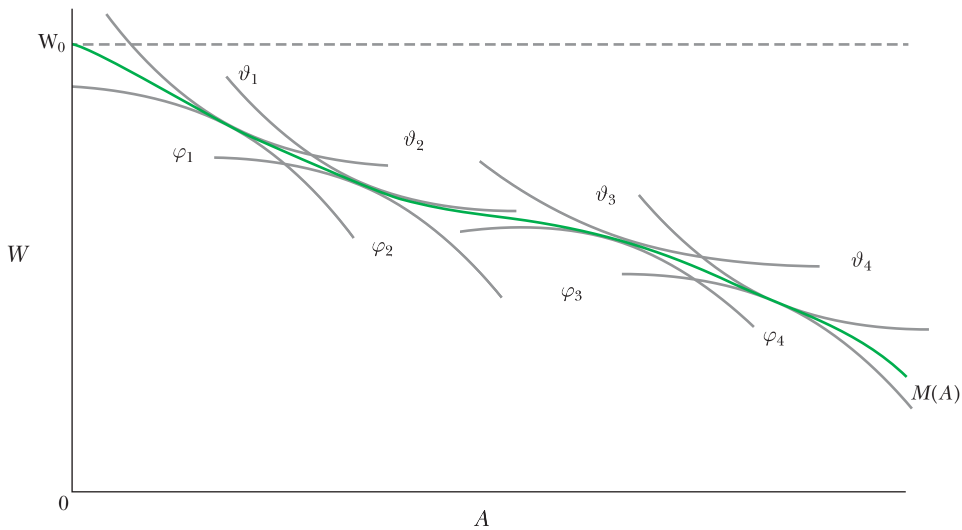
So far, the $M(A)$ function has been taken as given, but what determines this function? Answering this question requires considering labor demand and the employer side of the problem. The original Rosen (1986) model considers the case of firms with linear technology, where output x is produced for sale, labor L is a production input, and $\alpha_j(A)$ is a linear productivity parameter. The productivity parameter may differ across firms, denoted by j. In order to provide amenities, firms must give up productivity; that is, the cost of providing A can be thought of as foregone units of output x. As an example, consider workplace safety on a manufacturing assembly line: increasing safety may require slowing the pace of the assembly line, which decreases productivity. To choose the optimal level of amenities A, the firm equates the marginal benefit and marginal cost of providing A. The marginal benefit comes from firms paying lower wages when they offer A, and the marginal cost comes from foregone productivity. Firms differ in the technology for providing amenities; those with lower marginal costs choose to offer higher levels of A, all else equal.[2]

Firms face tradeoffs between offering higher wages or more amenities, and have isoprofit curves that describe the (W, A) pairs that hold profits fixed. Figure 2 depicts these isoprofit curves, denoted $\phi_1 - \phi_4$, for four firms that face different marginal costs of providing A. Firm ϕ_1 faces a high cost, so its isoprofit functions begin to decline steeply at low levels of A. Firms $\phi_2 - \phi_4$ have different technology functions that allow them to provide amenities at a relatively lower cost.

Workers' preferences can also be expressed by indifference curves over wages and amenities. For each worker, the optimal wage-amenity pair is the point along the $M(A)$ function that maximizes utility. As Figure 2 depicts, workers with preferences described by indifference curve θ_1 maximize utility by choosing the point on $M(A)$ that is tangent to their indifference curve. Workers with different preferences

[2] Although this assumption that amenity levels affect firm productivity is used in the original Rosen (1986) model, it is straightforward to change this assumption and consider the case in which A is costly but has no direct effect on productivity.

Figure 2
Sorting and Matching in the Labor Market



Source: Figure created by author.
Note: The horizontal axis shows amenity levels, A. The vertical axis shows wage levels, W. Four isoprofit curves, denoted $\phi_1 - \phi_4$, illustrate four firms that face different marginal costs of providing A. Four indifference curves, denoted $\theta_1 - \theta_4$, illustrate the preferences of four workers over tradeoffs between wages and amenities. The market equalizing differences curve, $M(A)$, is traced out by the set of tangency points between workers' indifference curves and firms' isoprofit functions.

sort along the $M(A)$ function according to their relative preferences for wages and amenities.

Putting both sides of the market together, the market equalizing differences curve, $M(A)$, is defined by the set of tangency points between workers' indifference curves and firms' isoprofit functions. In equilibrium, workers with different preferences for amenities are matched to firms with different marginal costs of providing amenities. The sorting and matching process that generates $M(A)$ is an essential component of the equilibrium. The shape of the $M(A)$ function depends on many factors, including the distribution of preferences, the distribution of firm technology parameters, output prices, the relative supply of workers at each level of A, and the relative demand for workers at each level of A. $M(A)$ is not generally a linear function.

In this literature, researchers have defined the idea of "compensating wage differentials" in various ways. I follow Rosen (1986) in defining the terms "compensating wage differential" and "equalizing wage differential" synonymously as the slope of the $M(A)$ market equalizing differences curve—that is, the rate at which market wages change as amenity levels change.

In this equilibrium, three classes of parameters are frequently objects of interest for empirical researchers and policymakers: (1) parameters that characterize workers' preferences and their marginal willingness to "pay" for amenities via lower wages; (2) parameters that characterize firms' costs of providing an amenity; and (3) parameters that characterize the $M(A)$ function or its slope, the market compensating wage differential. The empirical approaches to estimating these classes of parameters differ, and conflating these three conceptual objects has been a source of confusion for empirical interpretation and policy applications.

In particular, the $M(A)$ function is generally neither a representation of workers' preferences nor firms' technology—it arises from the sorting process that matches workers to firms. The market compensating wage differential is a local measure of preferences for the *marginal worker* whose indifference curve is tangent to the $M(A)$ function at a certain level of amenities. Among all workers who accept a job with a given amenity level, the marginal worker is the worker with the lowest willingness to pay. Therefore, the slope of the $M(A)$ curve provides little information about the preferences of an average worker. However, it does provide a lower bound on the preferences of inframarginal workers.

As an example, consider the case of jobs that require traveling. Some workers may prefer the opportunity to travel for work, while others may view work-related travel as disruptive to their lives.[3] The marginal worker might be someone with preferences somewhere in the middle. Because the $M(A)$ function is based on the preferences of only the marginal worker, the market compensating wage differential for jobs that require travel could potentially be zero even if most workers have strong preferences in one direction or the other.

In addition, the $M(A)$ function can shift for reasons unrelated to either the preferences of workers or the technology of firms. A shock to the output price of a good can move firms' isoprofit functions, shifting the profit-maximizing tradeoff between wages and amenities. For example, Charles et al. (2022) find that a 10 percent shock to output prices leads to a 1.5 percent increase in injury rates on average. Even if all workers' preferences remain the same, labor demand shifts may change how workers are sorted into jobs and thus shift the $M(A)$ function. With a single cross-sectional dataset, it is not typically possible to identify the full distribution of workers' preferences, or even *average* preferences, without imposing further assumptions, such as an assumption on the preference distribution (Ashenfelter 2006; Rosen 1986).[4]

Some special cases of the Rosen model can simplify estimation (Hwang, Mortensen, and Reed 1998; Rosen 1974). If firms have the same technology functions, workers with heterogeneous preferences sort along a single common isoprofit function. In this case, the market equalizing difference function is equivalent to the isoprofit function. Variation in labor supply can be used to estimate firm technology parameters. Similarly, if workers share common preferences, then changes in firm

[3] For related work on how workers sort differently across jobs after having children, see Hotz, Johansson, and Karimi (2018), who study shifts towards jobs with more family friendly amenities such hours flexibility.
[4] There are of course other approaches to estimating workers' preferences that do not rely on equilibrium labor market data. See Mas and Pallais (2017) for an example.

technology can be used to trace out the slope of indifference curves, identifying preferences.

A related scenario, an important special case for empirical researchers, occurs when firms share a common isoprofit function and purchase amenities in a competitive intermediate market instead of producing them. For example, firms may purchase health insurance for workers or contribute to retirement savings accounts. This scenario differs from the model depicted in the above graphs if providing amenities does not impact productivity and firms all face the same market price of amenities. Again, if firms share a common isoprofit function in a perfectly competitive labor market, workers with different preferences sort along this function and the isoprofit function is equivalent to $M(A)$. The $M(A)$ function may be linear if all firms face the same constant unit cost per worker for the amenity in a competitive labor market (and there are no tax-related distortions). In this case, the market-compensating wage differential is simply the market cost of the amenity, which may be observable to researchers. Of course, this theoretical result relies upon strong assumptions, including perfectly competitive labor and amenity markets, which may not hold in real-world empirical settings like markets for health insurance.

A Chronology of Empirical Methods, Estimation Challenges, and Limitations

I begin by presenting a cross-sectional approach taken in many earlier studies and then discuss adjustments that researchers made as panel data became more widely available. These panel-based approaches are designed to account for unobserved differences across workers and employers and imperfectly competitive labor markets. Finally, I discuss recent methods combining panel data with natural experimental designs involving exogenous amenities variation.

Cross-Sectional Wage Models

An extensive literature has sought to estimate compensating wage differentials by using a framework built on an ordinary least squares regression like:

$$W_i = X_i\beta + A_i\gamma + \varepsilon_i,$$

where W_i is the log wage of worker i; X_i is a vector of observed worker, job, and employer wage shifters, in addition to, ideally, any other market factors that may shift wages for reasons unrelated to amenities, such as time and location; and A_i is a vector of observed job amenities.[5] The aspiration behind this approach is to include in X_i all of the factors that systematically shift wages. For example, X_i should include worker skills or ability measures that affect worker productivity so that conditional

[5] For readers who wish to learn more about this cross-sectional literature, some helpful starting points include Viscusi and Aldy (2003), Duncan (1976), and Brown (1980).

on X_i workers are productively homogeneous. Under the assumption that X_i contains a sufficiently rich set of control variables, γ provides information about the average compensating wage differential for each amenity A. In practice, there are many challenges to implementing this approach. Here, I focus on a few of the most important and pervasive estimation and interpretation challenges, although a complete account is beyond the scope of this article.

In an example of this approach, Garen (1988) studies the relationship between log wages and the risk of injury on the job and estimates the amenity coefficient γ to be 0.0024. This result suggests that if a job increases a worker's annual risk of death by 1 in 100,000, then wages are on average higher by about 0.24 percent. For reference, a typical US manufacturing job has a fatal injury rate of about 3.5 deaths per 100,000 worker-years.[6] Therefore, the author's estimate of γ suggests that average manufacturing wages are about 0.84 percent higher than they would be if all fatal injury risks could be (hypothetically) eliminated.

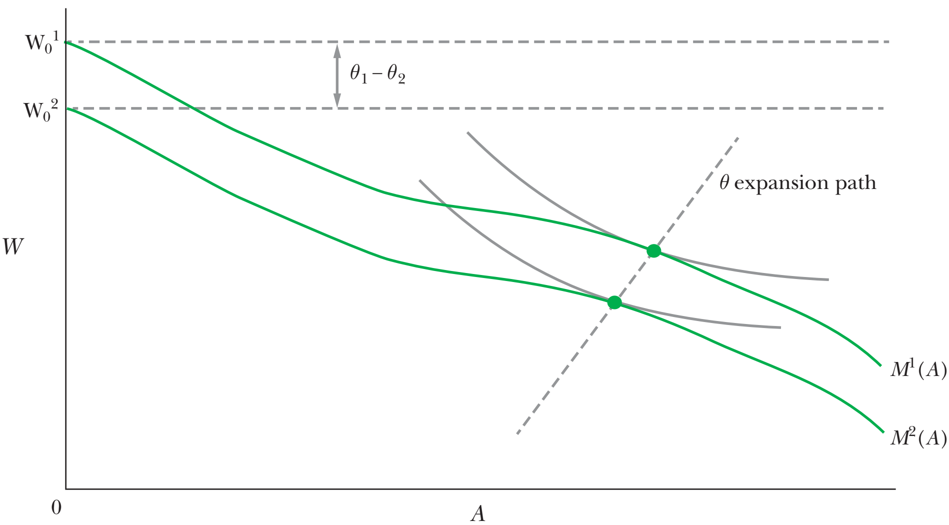
The basic regression version of the cross-sectional model assumes that the market-equalizing function is linear (with slope γ), thus ruling out the possibility of heterogeneity in compensating wage differentials. However, it is generally improbable in the Rosen model that all the tangency points between indifference curves and isoprofit functions will happen to lie on a straight line (Ekeland, Heckman, and Nesheim 2004). Thus, it is typically good practice to evaluate the linearity assumption relative to a more flexible function that may be nonlinear in A, although researchers frequently have limited statistical power to reject one functional form in favor of another.

Another challenge with the cross-sectional approach is that it is impractical to measure all of the dimensions of worker skills or abilities that influence wages and include them in X_i. Any unobserved components of ability are likely to cause endogeneity bias when estimating γ because high-earning workers tend to prefer jobs with more amenities, creating a positive correlation between ε and A.

In a classic example, Lucas (1977) estimates a cross-sectional wage model that controls for age, gender, schooling, and union membership and estimates compensating differentials for a set of amenities, including workplace hazards, whether the job requires repetitive tasks, is physically demanding, or requires supervising workers (among other amenities). Lucas finds that among workers with a high school degree, jobs that require highly repetitive tasks pay around 10–25 percent higher hourly wages. An instructive if perplexing finding, however, is that the coefficient on physically demanding jobs is negative—that is, it suggests that workers are willing to sacrifice wages to obtain physically demanding jobs. As Lucas (p. 218) explains: "[I]t is highly probable this effect may be explained by the omission of some ability, other than schooling, which is possessed by people in sedentary occupations." Although Lucas does not address this problem of unobserved ability and

[6] This estimate is based on the Census of Fatal Occupation Injuries, available at https://www.bls.gov/iif/fatal-injuries-tables.htm.

Figure 3
Unobserved Skills as a Source of Bias



Source: Figure created by author.
Note: The horizontal axis shows amenity levels, A. The vertical axis shows wage levels, W. There are two types of workers: high productivity (θ_1) or low productivity (θ_2). High-productivity workers can access jobs along the equalizing differences curve $M^1(A)$, while low-productivity workers can access jobs along the lower curve, $M^2(A)$. If the amenity A is a normal good, then the utility-maximizing choice of wages and amenities (W, A) will lie along an upward-sloping path as worker productivity rises.

omitted variables, his speculation that unobserved ability may explain counterintuitive findings in cross-sectional studies was echoed in subsequent literature.

To gain some intuition about such counterintuitive results, suppose the assumption that workers are productively homogeneous does not hold. Instead, there are two types of workers with either high productivity (θ_1) or low productivity (θ_2). Productivity is observed by firms but unobserved by the econometrician. As Figure 3 depicts, high-productivity workers can access jobs along the equalizing differences curve $M^1(A)$, while low-productivity workers can access jobs along the lower curve, $M^2(A)$. If the amenity A is a normal good, income effects will lead workers with high productivity to use some of their additional earnings potential to buy more A. Thus, as productivity increases (holding preferences fixed), the utility-maximizing choice of wages and amenities (W, A) will lie along an upward-sloping path as worker productivity rises. The same intuition extends to workers with different preferences—at each point along $M(A)$ there is an upward-sloping θ expansion path.

A researcher using data from this labor market faces the problem that there are two types of variation in the data—variation due to preference-based sorting along the negatively-sloped market equalizing differences curve and variation due to ability differences along the positively-sloped expansion path. In addition, worker productivity θ_i is correlated with amenities A_i under the assumption that amenities are normal goods. For example, if one runs a regression of log wages on job amenities only, without controlling for any differences in education or human capital, wages are usually positively correlated with amenities, because highly-skilled workers disproportionately hold good, high-paying jobs that offer many amenities.[7] Therefore, failing to control for productivity shifters in θ_i will introduce bias in the γ coefficient. If A is a desirable amenity, the bias is usually positive, and if A is an undesirable disamenity, the bias is usually negative.[8]

Unobserved productivity or ability is just one example of an omitted variable that may introduce bias in the cross-sectional model. It is important to note that the same intuition behind this source of bias extends broadly to many different forms of systematic variation in earnings besides worker productivity. Researchers must also consider many other sources of wage dispersion among workers that might be correlated with the amenities A_i and/or the controls in X_i. For example, in wage models with search frictions, more experienced workers often earn systematically higher wages. Because these workers have higher earnings potential, they may choose to purchase more amenities A, creating a correlation between experience, the degree of friction in the labor market, and amenities. For this reason, having a thorough model of wages that incorporates many sources of wage dispersion can help alleviate bias in the estimation of compensating wage differentials. In practice, however, no matter how detailed the available data, it is difficult in a cross-sectional model to completely account for all of the factors that may influence wages, including ability, unobserved amenities, labor market frictions, efficiency wages, rent-sharing, or discrimination, among others—all of which are potential sources of estimation bias.

In some cases, researchers have reason to believe that compensating differentials vary with observed characteristics of workers or jobs; for example, workers with young children may be highly averse to the risk of a severe injury. Researchers may wish to quantify the differences in compensating differentials across such groups. However, estimating heterogeneity in compensating differentials requires a strong word of caution, as Thaler and Rosen (1976) explained. In their study, the authors use cross-sectional data to estimate the relationship between wages and the risk of injury on the job using an ordinary least squares regression of the form:

$$W_i = X_i\beta + A_i\gamma_1 + A_i s_i \gamma_2 + \varepsilon_i,$$

where A_i is a measure of injury risk, s_i are worker characteristics that interact with injury risk to affect wages, and the remaining variables are defined above. An example of s_i that the authors consider is age. If a severe injury to a younger worker

[7] In some scenarios, the selection problem is more complex. See DeLeire, Khan, and Timmins (2013) for a model of compensating wage differentials with Roy-style sorting in which the returns to human capital differ by occupation.

[8] However, the direction of bias is not straightforward if the omitted productivity component θ_i is multidimensional, or if A is multidimensional.

causes a greater decline in lifetime future consumption, one may hypothesize that compensating wage differentials are larger for younger workers. Another possibility is that individual risk preferences may change with age for reasons other than lifetime future consumption.

This approach may appear to be a straightforward extension of the basic ordinary least squares cross-sectional specification; however, as Thaler and Rosen (1976) cautioned, the interpretation of estimates from specifications of this form is actually a "rather thorny issue." The reason is that in the Rosen model, there can only be one single market equalizing differences function, $M(A)$, for productively equivalent workers. However, an empirical specification that also includes the heterogeneity term $A_i s_i \gamma_2$ will shift the market equalizing differences function vertically for workers with different values of s_i. Therefore, an empirical model that includes shifters in the market equalizing differences function to capture preference heterogeneity is fundamentally inconsistent with the Rosen model.

However, including shifters for combinations of personal traits and amenities like $A_i s_i \gamma_2$ in an empirical model is not *always* problematic. If the researcher believes that workers with different values of s_i have different productivity, *and* the differences in productivity scale with the level of amenities, then including this term in the model is an appropriate—and, in general, necessary—way to account for productive heterogeneity. Just as the omission of any productivity shifter (θ_i) from the model can generally cause endogeneity bias, so too can the omission of factors that shift productivity in ways that depend on amenity levels. In Thaler and Rosen's (1976) model s_i captures observed factors that shift whether workers are able to be productive in the presence of risks—say, whether workers have "nerves of steel." If s_i includes age, the model imposes a specification assumption that older workers with more experience have a relative advantage in performing tasks in the presence of physical hazards, so that productivity differences by age scale with risk A_i. When estimating this type of specification, researchers should use caution to clearly separate how functional form assumptions related to the equalizing differences function differ from assumptions related to how the worker productivity component might interact with amenities.[9]

This point also raises the general question: how might researchers describe heterogeneity in compensating wage differentials? Instead of using a shifter term like $A_i s_i \gamma_2$, it is first necessary to relax the assumption that log wages change linearly with A_i and instead allow γ to vary with A_i. After estimating a model of this form, one can quantify group-level differences in compensating wage differentials by estimating the average marginal effects of changes in amenities on wages for different groups of workers. If workers with different characteristics sort to different locations on this nonlinear equalizing differences function, the average slope of the

[9] Hersch (2011) also considers a model similar to this in which the wage effects of a disamenity, workplace sexual harassment rates, operate through both a productivity channel and a compensating wage differential channel.

equalizing differences function may differ across groups, providing information about heterogeneity in compensating wage differentials.

Panel Models: Unobserved Skills and the Ability Bias Puzzle

To address unobserved differences in worker skills, researchers can use panel data that track the same workers over time and in this way control for some traits unobservable in standard data, such as ability, good work habits, or perseverance, as long as these traits remain constant. In a structural model, Hwang, Reed, and Hubbard (1992) showed that failing to correct for differences in worker productivity leads to significant bias in the estimation of compensating wage differentials. To reduce this bias, researchers have used panel-based approaches, such as the model estimated by Brown (1980), which is similar to:

$$W_{it} = X_{it}\beta + A_{it}\gamma + \theta_i + \varepsilon_{it},$$

where θ_i are fixed person effects that control for any persistent, static differences in wages across workers—for example, differences that might plausibly be attributable to unobserved ability or skills.

Including fixed person effects in the model changes the interpretation of the γ coefficient relative to the earlier ordinary least squares model. The fixed person effect, θ_i, absorbs the impact on wages of all the person-level characteristics, like human capital, but it also absorbs the average amenity levels of the jobs held by each person. The remaining variation in amenities not controlled by θ_i is the within-person variation over time, so this becomes the only portion of the variation in amenities used to identify γ in this approach.

What causes within-person variation in amenities? The answer can vary. Some amenities, such as the location of a job, may remain fixed over time for all jobs at a particular firm or establishment. For such amenities, the γ coefficient can only be identified by workers moving to a new job at a different location. Relying on job mobility as a source of identifying variation raises an additional set of concerns, which I discuss below, related to why workers change jobs and whether the factors that may have contributed to the job change decision are included in the wage model.

Other amenities, like whether the job is physically demanding, may also change if a worker switches occupations or tasks while remaining employed at the same firm. For these amenities, the γ coefficient on amenities can be identified by job changes across firms, within-firm changes in occupations, or a combination of both. Still other amenities, like the generosity of employer-provided health insurance benefits, may change over time even if the job and tasks performed remain the same. This type of amenity does not require any job mobility to identify γ, though job mobility may also contribute to identifying variation depending on the choice of model specification. Using a variation in amenities that is not tied to job changes can avoid one form of estimation challenge, which is related to the possibility that the error term in the wage model may influence the decision to change jobs, causing endogeneity bias.

In the panel model considered by Brown (1980), the set of amenities included whether the job requires repetitive tasks, working under stress, physical strength, working under bad physical conditions (say, extreme temperatures or physical hazards), and the risk of death on the job. The risk of death was measured at the occupation level, so changes in this risk came primarily from job changes across occupations. In the main specification, Brown finds that the compensating wage differential for the risk of death is positive and statistically significant when estimating a cross-sectional model, but the coefficient shrinks by 84 percent and becomes statistically insignificant when the same data are used to estimate a panel-based model that includes person effects.

Indeed, much (though not all[10]) of the literature that considers panel-based models finds that adding person effects to a wage model leads estimates of compensating wage differentials to shrink substantially and frequently become statistically insignificant (Kniesner et al. 2012; Viscusi and Aldy 2003). This pattern has been described as the "skills bias puzzle." The puzzle is that if job amenities are normal goods, then theory suggests that if A is a desirable amenity the bias caused by omitting θ_i is positive, and if A is an undesirable disamenity the bias is negative. However, empirical evidence has typically found the opposite.

What factors might explain the skills bias puzzle? First, the assumptions of perfect competition in the Rosen model may not hold, which will affect the estimation of compensating wage differentials, as I discuss in the next subsection. Or if workers lack information about amenity levels at the time they make job choices, then equilibrium compensating wage differentials may reflect the subjective beliefs of the marginal worker rather than objective measures of amenities. Second, in cases where some workers like A_i but others dislike it, researchers may not be able to specify a hypothesis about the direction of the marginal worker's preferences. This could potentially explain why panel-based approaches to corrected bias have unexpected impacts on estimates. One of the reasons why the literature has focused on the risk of death as an amenity is because hypotheses about the preferences of the marginal worker seem easier to justify.

When researchers are unsure about the sign of compensating wage differentials, one approach is to use data on voluntary job changes to test whether the upper and lower bounds on amenities for job changers share the same sign. Using this approach with panel data, Villanueva (2007) shows that if a worker voluntarily moves to a job with fewer amenities, the change in wages provides an upper bound on the willingness to pay for amenities. Similarly, a lower bound on preferences can be estimated among workers who voluntarily move to jobs with more amenities. For example, Villanueva (2007) finds that for workers who indicate that their job has a heavy workload, both the upper and lower bounds on γ are negative. For some other job characteristics, however, the evidence is inconclusive about the sign of compensating differentials.

[10] Two exceptions are Garen (1988) and Hwang, Reed, and Hubbard (1992).

A third explanation for the skills bias puzzle is that including person effects may not entirely control for productivity differences. If workers gain skills over time—for example, due to on-the-job training—this change in skills over time can be a source of omitted variable bias, violating the assumption that workers have conditionally homogeneous productivity. In addition to affecting wages, unmodeled changes in skills over time may be correlated with occupational promotions or job changes, which are an important source of identifying variation in amenities in some panel models.

Imperfect Competition

The early empirical work on compensating differentials, which often faced anomalous results like wrong-sided coefficients and the skills bias puzzle, prompted researchers to broaden their focus to other models of wage variation. Several studies moved away from the Rosen framework and instead combined compensating wage differentials and models of imperfect competition in labor markets by introducing features like search frictions that create wage dispersion, even among equally productive workers receiving the same amenities (Bonhomme and Jolivet 2009; Dey and Flinn 2005; Hwang, Mortensen, and Reed 1998; Lamadon, Mogstad, and Setzler 2022; Lang and Majumdar 2004; Sullivan and To 2013; Taber and Vejlin 2020).

Such studies often moved away from the types of wage regressions described above as well. One early alternative approach was to estimate preferences for amenities in imperfectly competitive labor markets by focusing on job separation rates, rather than equilibrium wages, as a dependent variable. The intuition was that how long workers remain at a job can be informative about how much a worker values the characteristics of that job (for more on this approach, see Gronberg and Reed 1994; Rosen 1986). Willingness to pay for job amenities can be estimated by dividing the effect of amenities on the probability of exiting a job by the effect of wages on the probability of exiting a job (and multiplying by –1).[11]

In an influential theoretical model by Hwang, Mortensen, and Reed (1998), workers search in a frictional labor market for jobs that offer different levels of utility from a combination of wages and nonwage amenities, where firms differ in their cost of providing amenities. Because firms that can more efficiently provide amenities at a lower cost can offer workers greater utility while still earning nonnegative profits, this leads to multiple equalizing differences functions, violating the basic structure of the Rosen equilibrium. In this scenario, even modest search frictions

[11] For example, consider an ordinary least squares regression in which the dependent variable is the probability a worker exits a job in a given year, and independent variables include the log wage and whether the job requires working in extreme conditions (say, very hot or cold temperatures). Suppose, hypothetically, the estimated coefficient on log wage is –0.8, and the coefficient on extreme job conditions is 0.2. The marginal willingness to accept a job with extreme conditions can be computed as $-1 \times 0.2/-0.8 = 0.25$, which implies that workers require 25 percent higher wages to accept a job with extreme conditions.

in the labor market cause very large bias in the estimation of workers' underlying preferences.

Imperfect competition may also help explain the skills bias puzzle. Suppose the labor market is imperfectly competitive, and three main components explain the variation in the observed data on wage and amenity pairs: (1) differences in preferences that lead workers to sort to different points along the equalizing differences function; (2) differences in worker ability that shift the equalizing differences function; and (3) heterogeneity in firm costs combined with search frictions that lead to shifts in the equalizing differences function. In this case, estimating the cross-sectional wage model described earlier will typically yield biased estimates of the γ coefficient on amenities, because the model omits the wage variation caused by the second and third components.

In addition, a panel data model with fixed worker effects will not necessarily reduce the total bias relative to the cross-sectional model. Remember that whereas the cross-sectional model uses across-worker variation caused by sorting along $M(A)$ to estimate the γ coefficient on amenities, the worker effects model uses only within-worker wage changes. In a panel data model controlling for worker effects θ_i, the variation in wages used to estimate the γ coefficient on amenities comes mainly from the third component—workers moving to jobs at firms that offer more utility because they face lower costs. However, this source of wage variation contains almost no information about the slope of the market equalizing differences function $M(A)$. When workers change jobs, they often move to jobs that offer both higher wages and more amenities. By isolating this component of the wage variation, within-worker estimates of compensating wage differentials may, in some cases, have even larger bias than cross-sectional estimates that omit worker effects (θ_i) as a control.

Matched Data Models

Rosen (1986) speculated 35 years ago that many estimation issues concerning compensating differentials would not be resolved until matched worker-firm data became available—that is, data that include linkages between each worker and each firm. Because matched employer-employee data, especially administrative censuses of jobs, contain information about how wages change when workers move between firms, they can be used to control for unobserved worker ability while also controlling for unobserved differences across firms, including unobserved job amenities or compensation policies that cause wages of similar workers to vary across firms. Matched data can also be used to relax the strong assumptions about perfect competition that are required in most cross-sectional or within-worker estimates of compensating differentials.

To gain intuition about how matched data allow estimation improvements, consider a framework that extends the panel model with worker effects so that it also includes fixed employer effects, as in:

$$W_{it} = X_{it}\beta + A_{it}\gamma + \theta_i + \psi_{J(i,t)} + \varepsilon_{it},$$

where the term $\psi_{J(i,t)}$ is a set of fixed employer effects. (The subscript $J_{(i,t)}$ can be read as employer J at which worker i is employed at time t.) The $\psi_{J(i,t)}$ term controls for static unobserved differences shared by all jobs in firm J that impact wages: for example, some firms systematically pay higher wages to reduce turnover rates. In the context of compensating differentials, the $\psi_{J(i,t)}$ term controls for compensating wage differentials associated with all of the unobserved amenities shared by jobs at firm J.[12]

This approach addresses a long-standing concern of researchers: it seems nearly impossible to quantify all of the characteristics that workers value across jobs. Including fixed employer effects reduces the need to observe every amenity, as long as the job characteristics are shared by workers at each firm and remain static over time. It is also possible to use matched data to control for unobserved factors that differ at other levels—for example $\psi_{J(i,t)}$ could be replaced with a firm-by-occupation term to control for unobserved factors that differ across occupation by firm pairs.

The conceptual goal behind this specification is to control for both unobserved worker ability and unobserved firm-level characteristics, like unobserved amenities. By including both sets of factors that can shift $M(A)$, the aim is to control for several key sources of wage dispersion among observably similar workers, so that the *conditional* equalizing differences function is analogous to the Rosen equilibrium and its slope is an estimate of compensating wage differentials. This model relaxes a large set of assumptions related to unobserved factors that drive wage variation across employers and jobs, although it still requires many assumptions about how equilibrium wages are determined.

Lavetti and Schmutte (2022) evaluate models of this form to estimate compensating wage differentials for the risk of death on the job. They show that including interactions between $\psi_{J(i,t)}$ and coarse occupation controls in the wage model appears to alleviate important sources of bias caused by unobserved differences in worker ability and unobserved employer amenities, while also relaxing the strong assumptions in the Rosen model that labor markets must be perfectly competitive. In a cross-sectional wage regression, they find that a 1 in a 1,000 increase in the annual risk of death on the job is associated with a 28 percent increase in wages. Adding worker effects to the model to control for ability causes the compensating wage differential to fall to just 3.7 percent, an 87 percent decline compared to the cross-sectional model. This pattern is consistent with the skills bias puzzle. However, using the matched structure of the data to control for firm-level differences in pay $\psi_{J(i,t)}$, the estimated compensating wage differential increases back to 17 percent. They develop a theoretical framework to show how empirical estimates from this wage model relate to the three classes of parameters in the Rosen model and propose empirical tests to assess whether parameter estimates can be interpreted as preferences or market compensating wage differentials.

[12] This model is similar to Abowd, Kramarz, and Margolis (1999), but it includes a job amenities term.

Discrimination, Exclusion, and Sorting

The Rosen model assumes a competitive market in which workers are free to access any job. Much of the literature that has relaxed the assumption of perfect competition has done so in a way that introduces nondiscriminatory frictions, like costs of job search. A smaller segment of the literature has considered a form of imperfect competition in which firms discriminate by excluding specific groups of workers from certain jobs or occupations, or by hesitating to hire them at all (Hsieh et al. 2019; Lang and Majumdar 2004). The US labor market has high levels of occupational segregation by race and gender: for example, Alonso-Villar, Del Río, and Gradín (2012) calculate the occupational Gini coefficient, which measures occupational inequality on a scale from zero (perfectly equal) to one (fully segregated), to be 0.34 by gender and 0.16 by race and ethnicity. If this occupational segregation is caused even partially by employers using demographic stereotypes to sort workers into occupations, this can change the nature of sorting in the labor market, altering the market equalizing differences function and compensating wage differentials.[13]

Note that discrimination can occur on either side of the labor market. Antos and Rosen (1975) study the case of discrimination on the labor supply side, in the case of teachers who have racially discriminatory preferences over the share of students at a school who are black or white. The authors estimate that white teachers required an additional $600 annually (in 1965 dollars) to move from an all-white school to an all-black school, while black teachers required an additional $200 annually to move in the opposite direction.

Discrimination and exclusion create some as-yet unresolved challenges for interpreting parameters from a Rosen-style model. In some social science disciplines, including social psychology, preferences are thought to be endogenously formed by social norms and stereotypes, as opposed to being exogenous primitives. From this perspective, it is difficult even to articulate conceptually what the Rosen equilibrium might look like in the absence of discrimination, including the absence of endogenous impacts on preferences (Bertrand 2020).

Quasi-random Variation in Amenities

Several recent studies have combined panel-based wage models with quasi-random variation in amenities to estimate the responsiveness of wages to amenity variation. For example, Lee and Taylor (2019) study the effects of randomized federal safety inspections of manufacturing plants carried out by the Occupational Safety and Health Administration. They find that random inspections that force plants to correct safety violations reduce plant-level fatality rates by 45 percent. In response to the safety improvement, workers' relative wages at inspected plants declined by 2–3 percent, consistent with the compensating wage differentials model.

[13] A different, though related type of scenario may occur if amenities differ across groups of workers. For example, Gruber (1994) studies the effects of mandatory changes in health insurance that increase coverage of maternity care, which leads to differential shifts in the wages of women relative to men.

In another example of quasi-random variation in amenities, Lavetti (2020) studies compensating wage differentials for the risk of death in Alaskan fisheries. This labor market displays seasonal variation in workplace hazards driven by weather patterns as well as variation from policy changes that improve safety. In addition, workers' employment contracts change frequently across fishing seasons. In this setting, it is possible to estimate compensating wage differentials while holding fixed the unobserved factors that may cause wage levels to differ across specific worker-firm pairs. In this setting, the basic cross-sectional wage model overstates compensating wage differentials by 90 percent relative to a model that accounts for unobserved job-specific heterogeneity and potentially endogenous sorting in the labor market.

Random or quasi-random variation in amenities can be useful for overcoming some of the challenges in estimating compensating wage differentials, though this type of variation does not resolve all of the issues discussed above. A benefit of random variation in amenities is that it allows researchers to estimate compensating wage differentials using changes in wages before and after the amenity variation. For example, if researchers lack rich data on worker human capital, it may still be possible to produce reliable estimates of compensating wage differentials as long as the impact of human capital on wage levels remains fixed before and after the amenity change.

In addition, random variation in amenities may change the requirements of the wage model so that, rather than requiring the error term to be uncorrelated with wage components, one only needs changes in the error term to be uncorrelated with *changes* in wage components before and after the randomization. This requirement may be easier to satisfy in some labor market settings, but it may not always hold. For example, imperfect competition that affects wage levels may also affect how responsive wages are to amenity changes, potentially biasing estimates of compensating wage differentials even with random amenity variation. Similarly, large amenity changes could lead workers to re-sort into different jobs, causing a form of endogenous variation in wage-amenity pairs in response to the random amenity variation. Despite these caveats, combining this type of amenity variation with an appropriately specified wage model is a promising direction for future progress in estimating compensating wage differentials.

Policy Applications

The Value of Statistical Life

One policy application of compensating differentials is to the "value of statistical life," which is a framework used to evaluate policies related to public safety or public health that involve the risk of death (Viscusi and Aldy 2003). Conceptually, the idea behind the value of statistical life is to imagine that many people each face a small risk of death and to determine how much money each person would be willing to pay to reduce everyone's average risk of death by a small amount such that on average one fewer person will die. The aggregate willingness to pay to prevent one expected death is the value of statistical life. For example, if 100,000 people are indifferent between accepting \$100 in additional annual earnings or a 1 in 100,000 increase in the annual risk of a fatal workplace accident, then the implied value of statistical life is \$10 million (\$100 $\times$ 100,000).

The Rosen framework has been used extensively to estimate workers' preferences for tradeoffs between wages and risks of death on the job—and the value of statistical life implied by these preferences. Many US federal regulatory agencies like the Environmental Protection Agency, the Food and Drug Administration, and the Federal Aviation Administration evaluate the costs and benefits of policies using estimates of the value of statistical life derived from labor market estimates of compensating wage differentials. For example, the US Environmental Protection Agency (2010) reviewed the literature on compensating wage differentials in labor markets and estimates of risk preferences derived from surveys where respondents were directly asked about their preferences. The agency currently places the value of saving a life at \$7.4 million (in 2006 dollars). When conducting a cost-benefit analysis of the Clean Air Act, the Environmental Protection Agency estimated that over 90 percent of the benefits came from the value of mortality reductions and were therefore based on estimates of the value of statistical life (Evans and Taylor 2020).

The labor-market setting is an appealing context to estimate the value of statistical life because it is a market with a relatively broad set of participants for whom it is common to face risk-wage tradeoffs. But based on the earlier discussion, some shortcomings of this approach become clear. Given that the value of statistical life is used to quantify society's preferences for risk reduction, consistency suggests that the value of statistical life should be calculated by aggregating workers' preferences—rather than using market compensating wage differentials. As discussed earlier, the $M(A)$ function and its slope are not a representation of workers' preferences: they are defined largely by the equilibrium sorting process that matches workers to firms.

In addition, the Rosen model identifies the preferences of the *marginal worker*. But many policies evaluated using the value of statistical life affect people who are not in the labor force (say, through safety regulations created by the Food and Drug Administration), so the risk preferences of nonworkers are not reflected in the estimates. Moreover, the marginal worker in any given job is the worker who values safety the least, which suggests that estimates of the value of statistical life derived from labor markets are likely to understate the average preferences of all workers in the labor force.

Given these limitations of the standard approaches to estimating compensating differentials, policymakers also consider a variety of alternative approaches to calculating the value of a statistical life, including using different datasets in labor market analyses as well as estimates of risk preferences derived from product markets, behaviors, or direct survey elicitation of preferences (Ashenfelter 2006; Cropper, Hammitt, and Robinson 2011).

Wage Gaps and Compensation Inequality

Earnings inequality has increased substantially in the United States and other advanced economies over the past half-century (for example, Hoffmann, Lee, and

Lemieux 2020.) Over the same time, the cost of employer-provided benefits, the largest component of which is health insurance in the United States, also increased enormously. In 2020, the worker with the median wage in the United States earned $19.15 per hour in earnings and their employers paid an additional $9.25 per hour, or 48 percent of wages, in total benefits (US Bureau of Labor Statistics 2020). These employer-provided benefits represent some of the amenities that workers might value, but additional amenities are excluded from this accounting cost, such as flexible hours and work-from-home options. Given the scale of nonwage benefits, inequality in total compensation—including the value to the worker of job amenities—may be very different than inequality in labor market earnings.

One approach to improving measures of compensation inequality would be to add compensating wage differentials to labor market earnings to calculate the market value of total compensation. For example, economists have estimated compensating wage differentials for health insurance, pension or retirement benefits, safety or workplace hazards, hours flexibility, job security, vacation or sick time, commuting time, and many other important amenities. Bell (2020) uses estimates of compensating wage differentials to construct inequality measures based on total compensation. Between black and white workers, the racial compensation gaps for workers are slightly larger than wage gaps; between males and females, the gender compensation gap may be up to two-thirds smaller than the gender wage gap. One obvious caveat to this approach is that researchers are limited to observed amenities.

In contrast to this bottom-up approach, Taber and Vejlin (2020) and Sorkin (2018) develop top-down measures of the total value of the bundle of unobserved amenities. The intuition is that if a worker voluntarily exits a job and moves to a new job that pays lower wages, the total value of amenities at that new job must exceed the value of amenities at the original job. Connecting this type of revealed preference information for all job changes using matched employer-employee data from a large set of firms, Taber and Vejlin (2020) estimate that compensating wage differentials explain a larger share of inequality in the utility of jobs than wages do. This result may still understate the full importance of compensating wage differentials because the net value of a bundle of job amenities often contains partially offsetting positive and negative wage differentials for good and bad job characteristics.

The reasons behind the dispersion of wages may affect how policymakers view these differences. If a large share of earnings inequality is caused by underlying differences in human capital, policymakers might wish to assess education and job training policies. If labor market frictions explained a large share of earnings differences, policymakers might wish to evaluate labor market institutions and regulations. However, if compensating wage differentials contribute meaningfully to earnings dispersion across groups, this variation may be less concerning for policymakers if it reflects differences in the preferences of workers who sort into jobs that differ in wages because they offer different levels of amenities that workers value.

Indeed, one could make a case that a measure of compensation that removes the effects of compensating wage differentials is *preferable* to observed income when assessing inequality. If a worker chooses to accept a lower wage and implicitly "purchases" a flexible schedule (or even a company car), should this purchase be viewed differently from an inequality perspective than the purchase of other goods or services? In this sense, netting out compensating differentials from observed earnings is consistent with a measure of earnings inequality that is agnostic regarding how people spend their total earnings. Of course, constructing such a measure is challenging and requires addressing the many estimation issues I discuss above.

Conclusion

The compensating wage differentials model is widely regarded as one of the core models of wage determination in labor economics. Empirically, compensating wage differentials are large enough to matter when economists consider topics like wage dispersion, total compensation, choices of occupations, patterns of discrimination, and even issues that may at first glance seem unrelated, like how policymakers might place a monetary value on the benefits of risk-reducing regulations. However, the literature on how to model and estimate compensating wage differentials has had to address many challenges, and has evolved considerably in recent decades.

On one side, introducing imperfect competition led the literature on compensating wage differentials to splinter as it moved away from the competitive Rosen model and considered alternative models of frictional search, bilateral bargaining, sorting, and rent-sharing. Much of this literature took a pessimistic view on whether researchers would be able to disentangle the influences of imperfect competition from compensating wage differentials.

On the other side, the expansion of matched employer-employee data, and the use of quasi-random variation in job amenities, has kindled some new optimism as researchers developed novel approaches to estimating compensating wage differentials using models that relax the assumptions of perfect competition. In particular, considering the net effect on wages of an aggregate bundle of unobserved amenities has provided new insights into the importance of the compensating wage differentials model for earnings and compensation inequality and allows researchers to consider questions that had never previously been studied. But these approaches are still nascent, and there is much to be learned.

References

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67 (2): 251–333.

Alonso-Villar, Olga, Coral Del Río, and Carlos Gradín. 2012. "The Extent of Occupational Segregation in the United States: Differences by Race, Ethnicity, and Gender." *Industrial Relations* 51 (2): 179–212.

Antos, Joseph R., and Sherwin Rosen. 1975. "Discrimination in the Market for Public School Teachers." *Journal of Econometrics* 3 (2): 123–50.

Ashenfelter, Orley. 2006. "Measuring the Value of a Statistical Life: Problems and Prospects." *Economic Journal* 116 (510): C10–23.

Bell, Alex. 2020. "Job Amenities and Earnings Inequality." Unpublished.

Bertrand, Marianne. 2020. "Gender in the Twenty-First Century." *AEA Papers and Proceedings* 110: 1–24.

Bonhomme, Stephanie, and Gregory Jolivet. 2009. "The Pervasive Absence of Compensating Differentials." *Journal of Applied Econometrics* 24 (5): 763–95.

Brown, Charles. 1980. "Equalizing Differences in the Labor Market." *Quarterly Journal of Economics* 94 (1): 113–34.

Charles, Kerwin Kofi, Matthew S. Johnson, Melvin Stephens Jr., and Do Q. Lee. 2022. "Demand Conditions and Worker Safety: Evidence from Price Shocks in Mining." *Journal of Labor Economics* 40 (1): 47–94.

Cropper, Maureen, James K. Hammitt, and Lisa A. Robinson. 2011. "Valuing Mortality Risk Reductions: Progress and Challenges." *Annual Review of Resource Economics* 3 (1): 313–36.

DeLeire, Thomas, Shakeeb Khan, and Christopher Timmins. 2013. "Roy Model Sorting and Nonrandom Selection in the Valuation of a Statistical Life." *International Economic Review* 54 (1): 279–306.

Dey, Matthew S., and Christopher J. Flinn. 2005. "An Equilibrium Model of Health Insurance Provision and Wage Determination." *Econometrica* 73 (2): 571–627.

Duncan, Greg J. 1976. "Earnings Functions and Nonpecuniary Benefits." *Journal of Human Resources* 11 (4): 462–83.

Ekeland, Ivar, James J. Heckman, and Lars Nesheim. 2004. "Identification and Estimation of Hedonic Models." *Journal of Political Economy* 112 (S1): S60–109.

Evans, Mary F., and Laura O. Taylor. 2020. "Using Revealed Preference Methods to Estimate the Value of Reduced Mortality Risk: Best Practice Recommendations for the Hedonic Wage Model." *Review of Environmental Economics and Policy* 14 (2): 282–301.

Flood, Sarah, Miriam King, R. Rodgers, Steven Ruggles, R. Warren, and M. Westberry. 2022. "Integrated Public Use Microdata Series, Current Population Survey: Version 10.0 [dataset]." IPUMS. https://doi.org/10.18128/D030.V10.0 (accessed December 14, 2022).

Garen, John. 1988. "Compensating Wage Differentials and the Endogeneity of Job Riskiness." *Review of Economics and Statistics* 70 (1): 9–16.

Gronberg, Timothy J., and W. Robert Reed. 1994. "Estimating Workers' Marginal Willingness to Pay for Job Attributes Using Duration Data." *Journal of Human Resources* 29 (3): 911–31.

Gruber, Jonathan. 1994. "The Incidence of Mandated Maternity Benefits." *American Economic Review* 84 (3): 622–41.

Hersch, Joni. 2011. "Compensating Differentials for Sexual Harassment." *American Economic Review* 101 (3): 630–34.

Hoffmann, Florian, David S. Lee, and Thomas Lemieux. 2020. "Growing Income Inequality in the United States and Other Advanced Economies." *Journal of Economic Perspectives* 34 (4): 52–78.

Hotz, V. Joseph, Per Johansson, and Arizo Karimi. 2018. "Parenthood, Family Friendly Firms, and the Gender Gaps in Early Work Careers." NBER Working Paper 24173.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. 2019. "The Allocation of Talent and U.S. Economic Growth." *Econometrica* 87 (5): 1439–74.

Hwang, Hae-shin, Dale T. Mortensen, and W. Robert Reed. 1998. "Hedonic Wages and Labor Market Search." *Journal of Labor Economics* 16 (4): 815–47.

Hwang, Hae-shin, W. Robert Reed, and Carlton Hubbard. 1992. "Compensating Wage Differentials and Unobserved Productivity." *Journal of Political Economy* 100 (4): 835–58.

Kniesner, Thomas J., W. Kip Viscusi, Christopher Woock, and James P. Ziliak. 2012. "The Value of Statistical Life: Evidence from Panel Data." *Review of Economics and Statistics* 94 (1): 74–87.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler. 2022. "Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market." *American Economic Review* 112 (1): 169–212.

Lang, Kevin, and Sumon Majumdar. 2004. "The Pricing of Job Characteristics When Markets Do Not Clear: Theory and Policy Implications." *International Economic Review* 45 (4): 1111–28.

Lavetti, Kurt. 2020. "The Estimation of Compensating Wage Differentials: Lessons from the Deadliest Catch." *Journal of Business and Economic Statistics* 38 (1): 165–82.

Lavetti, Kurt, and Ian M. Schmutte. 2022. "Estimating Compensating Wage Differentials with Endogenous

Job Mobility." Unpublished.
Lee, Jonathan M., and Laura O. Taylor. 2019. "Randomized Safety Inspections and Risk Exposure on the Job: Quasi-experimental Estimates of the Value of a Statistical Life." *American Economic Journal: Economic Policy* 11 (4): 350–74.
Lucas, Robert E. B. 1977. "Hedonic Wage Equations and Psychic Wages in the Returns to Schooling." *American Economic Review* 67 (4): 549–58.
Mas, Alexandre, and Amanda Pallais. 2017. "Valuing Alternative Work Arrangements." *American Economic Review* 107 (12): 3722–59.
Rosen, Sherwin. 1974. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82 (1): 34–55.
Rosen, Sherwin. 1986. "The Theory of Equalizing Differences." In *Handbook of Labor Economics*, Vol. 1, edited by Orley C. Ashenfelter and Richard Layard, 641–92. Amsterdam: North Holland.
Smith, Adam. 1776. *The Wealth of Nations.* New York: Modern Library, 2000.
Sorkin, Isaac. 2018. "Ranking Firms Using Revealed Preference." *Quarterly Journal of Economics* 133 (3): 1331–93.
Sullivan, Paul, and Ted To. 2013. "Job Dispersion and Compensating Wage Differentials." US Bureau of Labor Statistics Working Paper 469.
Taber, Christopher, and Rune Vejlin. 2020. "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market." *Econometrica* 88 (3): 1031–69.
Thaler, Richard, and Sherwin Rosen. 1976. "The Value of Saving a Life: Evidence from the Labor Market." In *Household Production and Consumption*, edited by Nestor E. Terleckyj, 265–302. New York: NBER.
US Bureau of Labor Statistics. 2020. "Employer Costs for Employee Compensation—March 2020." https://www.bls.gov/news.release/archives/ecec_06182020.pdf (accessed March 7, 2023).
US Environmental Protection Agency. 2010. "Valuing Mortality Risk Reductions for Environmental Policy: A White Paper." https://www.epa.gov/environmental-economics/valuing-mortality-risk-reductions-environmental-policy-white-paper-2010.
Villanueva, Ernesto. 2007. "Estimating Compensating Wage Differentials Using Voluntary Job Changes: Evidence from Germany." *Industrial and Labor Relations Review* 60 (4): 544–61.
Viscusi, W. Kip, and Joseph E. Aldy. 2003. "The Value of a Statistical Life: A Critical Review of Market Estimates throughout the World." *Journal of Risk and Uncertainty* 27 (1): 5–76.

Journal of Economic Perspectives—Volume 37, Number 3—Summer 2023—Pages 213–232

What Can Historically Black Colleges and Universities Teach about Improving Higher Education Outcomes for Black Students?

Gregory N. Price and Angelino C. G. Viceisza

Historically Black colleges and universities, often referred to as HBCUs, are institutions that were founded and developed in an environment of legal segregation prior to the Civil Rights Act of 1964 with the principal mission of educating Black Americans. The National Center for Education Statistics (2022, Table 313.10) listed 99 such institutions located in 19 states, the District of Columbia, and the US Virgin Islands. Total enrollment at these institutions in 2021 was 287,000 students.

In their most important function—enrolling and graduating college students—historically Black colleges and universities punch significantly above their weight (Morse, Sakano, and Price 1996; Sharpe 2016; Broady, Todd, and Booth-Bell 2017). Together, these institutions enroll about 1.5 percent of all college students, with about three-quarters of that enrollment being Black students. However, this accounts for 9 percent of all Black college students nationally and for 13 percent of all Black college graduates in 2021 (National Center for Education Statistics 2023). As we will point out later in this paper, nearly one-quarter of all Black PhDs in science-related fields graduated from these schools. Of course, these patterns vary across states (for discussion, Saunders and Nagle 2019). In Florida, historically Black colleges and universities enroll 9 percent of all Black undergraduates and award 18 percent of all bachelor's degrees to Black college graduates. In Louisiana, these institutions enroll 38 percent of all Black students–and produce a similar number of college

■ *Gregory N. Price is Professor of Economics, University of New Orleans, New Orleans, Louisiana. Angelino C. G. Viceisza is Associate Professor of Economics, Spelman College, Atlanta, Georgia. Their emails are gnprice@uno.edu and aviceisz@spelman.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.37.3.213.

graduates. In Virginia, they enroll 29 percent of the state's Black college students and 32 percent of its Black college graduates.

Historically Black colleges and universities have a long tradition of reaching out to the disadvantaged, including those who are first-generation college students or who come with low socioeconomic status. Thus, taken as a group, the students at these schools are more likely than Black college students at other colleges and universities to have lower academic qualifications (for example, as measured in terms like high school grade point average and test scores) and more likely to be from socioeconomically disadvantaged families (for example, as measured by receiving Pell grants). Nathanson, Samayoa, and Gasman (2019) report that almost 24 percent of students at historically Black colleges and universities are low-income, three times the rate at primarily white institutions. Of course, looking at the averages for all students of historically Black colleges and universities as a group does not take the heterogeneity of these institutions into account.

Table 1 shows some 2021 characteristics for two types of four-year historically Black colleges and universities—first, the ten largest by enrollment, and second, some smaller historically Black institutions. Of the 16 institutions listed, nine are public. Enrollment ranges from 2,173 to 11,207, with six-year graduation rates varying from 21.7 percent to 75.1 percent. As mentioned previously, a significant proportion of students at these institutions come from disadvantaged backgrounds, as ten of the institutions listed have Pell percentages over 50 percent.

In this essay, we focus on two main issues. We start by examining how Black students perform across historically Black colleges and universities compared with other institutions along a relatively broad range of outcomes. We then consider some potential causes of these differences. In so doing, we provide a glimpse into the "secret sauce" that allows these historically Black schools to start with students who are below-average on measures of preparedness for college and end up with above-average results. We conclude with potential implications for policymakers and university administrators. We also emphasize that compared to other colleges and universities, historically Black colleges and universities have been under-resourced, both historically during the time of legal segregation and more recently as well.

Better for Black Students?

College-Level Outcomes

The graduation rates for Black students at historically Black colleges and universities is about 32 percent; for Black students at other institutions, the graduation rate is about 44 percent (Gordon et al. 2021). However, as noted already, historically Black institutions serve a distinctive group. Thus, evaluating the performance of Black students at historically Black institutions turns on how one adjusts for such differences.

For example, Gordon et al. (2021) adjust graduation rates using both parametric and "coarsened exact matching" methods (looking at schools with similar

Table 1
2021 Characteristics for Selected Four-Year Historically Black Colleges and Universities

Institution	*State*	*Type*	*Enrollment*	*Pell percent*	*Six-year Black graduation rate*
North Carolina A&T State University	NC	Public	11,207	57	51.8%
Florida Agricultural and Mechanical University	FL	Public	9,780	61	53.2%
Howard University	DC	Private	9,069	44	64.2%
Prairie View A&M University	TX	Public	9,011	65	35.5%
Texas Southern University	TX	Public	8,933	65	21.7%
North Carolina Central University	NC	Public	7,541	61	49.8%
Morgan State University	MD	Public	7,286	53	43.7%
Tennessee State University	TN	Public	6,875	55	29.6%
Jackson State University	MS	Public	6,568	70	37.1%
Southern University and A & M College	LA	Public	6,283	68	40.9%
Hampton University	VA	Private	4,510	36	59.8%
Tuskegee University	AL	Private	3,276	21	52.0%
Xavier University of Louisiana	LA	Private	3,133	21	51.4%
Morehouse College	GA	Private	2,296	48	55.0%
Spelman College	GA	Private	2,261	46	75.1%
Claflin University	SC	Private	2,173	71	50.7%

Source: This table is based on publicly available data at https://www.msidata.org.
Note: This table comprises four-year historically Black institutions with the top-ten enrollment as well as select smaller HBCUs; specifically, Hampton University, Tuskegee University, Xavier University of Louisiana, Morehouse College, Spelman College, and Claflin University.

characteristics) to compare institutional and individual factors. Institutional factors include the size of the institution, selectivity, finances, quality of instruction, setting, culture and atmosphere, alumni/legacy connections, and others. Individual student factors include test scores and socioeconomic status. They find that "African American students attending HBCUs are up to 33 percent more likely to graduate than African American students attending a similar non-HBCU." In an analysis with more detailed individual data, Hill, Kurban, and Swinton (Forthcoming) find that for Black students with intermediate test scores, attending a historically Black institution is associated with an increase of 13–14 percent in the graduation rate; for Black students with low test scores, the graduation rate doubles. In addition, when comparing Black students who graduate, the historically Black institutions were more likely to produce a science, technology, engineering, and mathematics graduate.

Several studies have found that differences in graduation and retention rates between historically Black colleges and universities and other institutions are mostly driven by selection (for other examples, see Richards and Awokoya 2010; de Zeeuw, Fazili, Hotchkiss 2021; Wilson 2007). There is some evidence that historically Black colleges and universities have a comparative advantage in economic education. Simkins and Allen (2000) gave the Test of Understanding College Economics (TUCE III) before and after introductory courses in economics to students at two regional state institutions of about the same size in the same city: the historically Black North Carolina A&T State University and the traditionally white University of North Carolina at Greensboro. They found that the students at North Carolina A&T scored lower on the pretest before the course began, but that scores were equivalent across the two schools after the end of the class.

Graduate School and the Academic Pipeline

Almost one-quarter (23.2 percent) of Black graduates who earned a doctorate in science and engineering between 2015 and 2019 earned their bachelor's degree from a historically Black college or university (National Science Foundation 2021). Historically Black institutions produced particularly large shares of baccalaureate recipients who later earned doctoral degrees in agricultural sciences and natural resources (almost 50 percent of all Black graduates who earned such a degree), computer and information sciences (over 30 percent), and mathematics and statistics (over 30 percent). The top producers were Howard University and Spelman College.

To get a more general picture beyond science and engineering, we turn to the Baccalaureate and Beyond dataset, a nationally representative longitudinal study of students who completed the requirements for a bachelor's degree in a given academic year, specifically 1992–1993 (cohort 1), 2000–2001 (cohort 2), 2007–2008 (cohort 3), and 2015–2016 (cohort 4).[1] Table 2 compares the proportion of students from historically Black colleges and universities who applied to graduate school ten years after completing college with Black graduates of other institutions. While the overall levels for the 1993/2003 cohort were higher than those for the 2008/2018 cohort, students from historically Black colleges and universities were about as likely as Black students at other institutions to have applied to graduate school ten years after college regardless of the cohort. Table 2 presents unadjusted averages; thus, it must be viewed as a success story that historically Black colleges and universities started with a higher share of socioeconomically and academically disadvantaged students, yet ended up with a similar share attending graduate school.

These statistics have implications for diversity, particularly in academia, but also in jobs after graduation and for knowledge production more broadly. Here, we focus on evidence on these themes from the field of economics. Price and Sharpe (2020) find that PhD-granting departments of economics with more Black faculty members tend to produce more research on topics related to race; thus, they argue that the low numbers of Black economists in these departments has constrained the production of economics knowledge that would reduce racial inequality and improve the

[1] The National Postsecondary Student Aid Study serves as the base year for each cohort. The Baccalaureate and Beyond data follow graduating seniors one, four, and ten years after completing their bachelor's degree. Ten-year data are only available for cohorts 1992–1993 and 2007–2008, which are typically referred to as 1993/2003 and 2008/2018. All statistics from these data are generated using DATALAB, https://bit.ly/3bvFvJO, as shown in Price and Viceisza (2023).

Table 2

A Selection of Average Black Student Outcomes across Historically Black Institutions and Non-HBCUs

Outcome	*HBCU*	*Non-HBCU*
Applied to Graduate School		
Cohort: 1993/2003	0.39	0.40
(Standard error)	(0.07)	(0.03)
Cohort: 2008/2018	0.27	0.28
(Standard error)	(0.04)	(0.02)
Annual Salary		
Cohort: 1993/2003	$47,072	$52,449
(Standard error)	($4,430)	($1,665)
Cohort: 2008/2018	$58,708	$59,527
(Standard error)	($3,357)	($1,787)
Pay Satisfaction		
Cohort: 1993/2003	0.35	0.60
(Standard error)	(0.04)	(0.04)
Cohort: 2008/2018	54.40	56.92
(Standard error)	(6.49)	(2.62)
Job Security Satisfaction		
Cohort: 1993/2003	0.69	0.80
(Standard error)	(0.05)	(0.03)
Cohort: 2008/2018	72.12	70.66
(Standard error)	(5.64)	(2.33)
Ever Voted		
Cohort: 1993/2003	0.88	0.85
(Standard error)	(0.04)	(0.03)
Cohort: 2008/2018	0.79	0.80
(Standard error)	(0.04)	(0.02)

Source: All statistics from these data are generated using DATALAB, https://bit.ly/3bvFvJO, as shown in Price and Viceisza (2023).
Note: This table presents averages for Black graduates from historically Black institutions and Black graduates from non-HBCUs. No other factors are controlled for.

living standards of Black Americans. Francis, Hardy, and Jones (2022) explore the contributions of Black economists to research on major economic and social policy problems in the United States by focusing on topics in education, poverty and economic mobility, and public finance. Of the more than 260 articles they cite, 11 percent include economists on faculties at historically Black colleges and universities. Agesa, Granger, and Price (2000) find that while economics departments at historically Black colleges and universities produced less research output relative to other economics departments, research output was equally effective in leading undergraduate economics majors to pursue economics doctorates at both types of institutions—which also suggests that support for research at historically Black institutions would likely lead to more Black PhD economists.[2]

Compensation and Satisfaction

We now turn to two labor market outcome measures from the Baccalaureate and Beyond dataset: salary and career satisfaction. Comparing Black graduates across all majors in terms of annual salary ten years after graduation, graduates of historically Black colleges and universities did worse relative to other institutions in the 1993/2003 cohort, but seem to have caught up in the 2008/2018 cohort as shown in Table 2.

These findings are consistent with, and in some sense reconcile, those of prior studies. Studies looking at data for the 1970s find that attending a historically Black institution provided a wage premium for Black college students at the time compared to Black students attending other institutions, after adjusting for individual factors. For example, Constantine (1995) finds this result using data from the National Longitudinal Survey of the Class of 1972. Price, Spriggs, and Swinton (2011) use data from the National Survey of Black Americans, conducted in four waves between 1979–1980 and 1992, and find benefits to both labor market and psychological outcomes from attending a historically Black institution. Similarly, Fryer and Greenstone (2010) adjust for family background, high school academic achievement, and statistical methods to correct for unobserved personal difference and find a wage premium from attending a historically Black college or university in the 1970s using nationally representative data. However, using the same methods, Fryer and Greenstone find a wage *penalty* for attending a historically Black institution in the 1990s. This is similar to our finding in Table 2. Wood and Palmer (2017), using data from the 1993/2003 Baccalaureate and Beyond, find that historically Black institutions are equivalent to other higher education institutions in terms of human capital formation to obtain comparable labor market outcomes.

Studies using more recent data tend to find a wage premium for Black students attending historically Black colleges and universities. For example, Elu et al. (2019) use data from the 2015 US Department of Education College Scorecard, which linked together data on characteristics of colleges and their students from the Integrated Postsecondary Education Data System, the National Student Loan Data System, federal earnings and tax records, and other sources. They find that after adjusting for factors like returns to different college majors and the urban wage premium, there is a long-run wage premium for a Black student attending a historically Black college or university.

Table 2 also compares two measures of career satisfaction: pay/compensation satisfaction and job security satisfaction (with the exact wording of the questions

[2] This finding is consistent with Biasi and Ma (2022) who find that (1) more selective and better funded schools, and those enrolling socioeconomically advantaged students, teach more frontier knowledge; and (2) students from these schools are more likely to complete a doctoral degree, produce more patents, and earn more after graduation.

varying slightly over time). For the 1993/2003 cohort, Black graduates of the historically Black institutions do worse on all measures than Black graduates of other institutions, which appears consistent with the salary findings reported previously. For the 2008/2018 cohort, Black graduates of historically Black institutions express more satisfaction with regard to job security than Black graduates of other institutions.

Thriving and Social Mobility

A Gallup (2015) poll conducted in 2014 and 2015 reports that among Black college graduates, those who attended a historically Black institution are more likely than others to agree strongly with the statement that they are thriving in purpose well-being (61 percent versus 40 percent) and financial well-being (40 percent versus 29 percent). Digging deeper into the survey data, these differences reflect a higher likelihood of the graduates of historically Black institutions expressing liking what they do each day, being motivated to achieve goals, and effectively managing their economic life to reduce stress and increase security.

Several studies have made use of the publicly available Opportunity Insights dataset, generated by government researchers working with Harvard researchers, to study the socioeconomic mobility of graduates from different types of institutions. These data are built on aggregate intergenerational (parent linked to child) income information from the Internal Revenue Service (IRS) linked with college attendance information from the NCES Integrated Postsecondary Education Data System (IPEDS). The IRS data link parents and their children's reported income earnings for children born from 1980–1991. Parents' household earnings are calculated as the five-year average when the child was age 15–19. Children's earnings are calculated as wage and self-employment earnings in 2014, when children were in their early to mid-30s.

Using these data, Nathenson, Samayoa, and Gasman (2019) find that historically Black colleges and universities enroll far more low-income students than predominantly white institutions, and more students experience upward mobility at historically Black rather than predominantly white institutions. They compare 50 historically Black institutions with 115 primarily white institutions, yielding an analytic sample of 165 institutions. For example, their calculations show that two-thirds of low-income students at historically Black colleges and universities end up in at least the middle class. In their estimate, Xavier University of Louisiana, Dillard University, and Tuskegee University are among the historically Black institutions doing a particularly good job at fostering upward mobility for their students.

Hammond, Owens, and Gulko (2021) also use this Opportunity Insights dataset. They compare 50 historically Black institutions with a broader group of 1,235 other colleges and universities. They measure an institution's "access" as the percentage of students whose parent or guardian had low income levels, and an institution's "success" as the share of students from low-income households who attain a higher income category. They multiply access and success to obtain a measure of how well an institution promotes mobility. By this measure, 52 percent of all institutions above the 95th percentile of mobility are historically Black colleges and universities. In part, this outcome arises because the historically Black institutions have an access rate twice the national average and over five times that of what they categorize as "Ivy Plus" institutions. The ten historically Black colleges and universities with the highest mobility rates are: Claflin University, Rust College, Grambling State University, Florida Memorial University, Southern University at New Orleans, Jackson State University, Elizabeth City State University, Mississippi Valley State University, South Carolina State University, and Southern University and A&M College at Baton Rouge.

Itzkowitz (2022) also uses the Opportunity Insights data to create his own economic mobility index. For the index, he calculates a "Price-to-Earnings Premium"—basically, how long it takes a low-income student to recoup the costs of their education based on the premium they earn by attending a given institution—and multiply their comparative rank by the percentage of Pell Grant recipients that each institution enrolls. He finds that Hispanic-serving institutions concentrated in California, Texas, and New York provide the most economic mobility by this measure, but historically Black institutions also score high. By this measure, the top ten historically Black institutions for social mobility are: Elizabeth City State University, Xavier University of Louisiana, North Carolina A&T State University, Fayetteville State University, Florida A&M University, Bowie State University, Prairie View A&M University, Tennessee State University, Winston-Salem State University, and Wilberforce University.

Voting and Health

One proxy for civic engagement is whether someone has ever voted. As shown in Table 2, Black graduates of historically Black colleges and universities are more civically engaged by this measure than Black graduates of other colleges and universities, although the difference seems to have decreased over time.

Attending a historically Black college or university seems to have a positive impact on the physical and mental health of Black students. Colen, Pinchak, and Barnett (2021) assess the odds of metabolic syndrome among Black college attendees age 24–34. Specifically, they assess whether respondents are above a high-risk threshold on at least three of five common biomarkers of cardiovascular health: blood pressure, waist circumference, triglycerides, high-density lipoprotein, and glycated hemoglobin levels. They find that enrollment in a historically Black institution is associated with a 35 percent reduction in such odds compared with those who attend other colleges and universities. Moreover, those who attend historically Black institutions and who grew up in more segregated environments experience the greatest reductions in the likelihood of developing metabolic syndrome. This finding is consistent with the Gallup (2015) evidence mentioned earlier.

With regard to psychological well-being, the Price, Spriggs, and Swinton (2011) study mentioned earlier using data from the National Survey of Black Americans found that Black students who attended a historically Black institution scored more highly on self-esteem and Black identity. In a sample of 171 students at historically Black institutions, Braby, Holcomb, and Leonhard (2022) found that those with high ethnic identification and resilience reported lower alcohol and substance use, as well as minimal mental health distress. During the recent and ongoing COVID-19 pandemic, Huang, Li, and Hsu (2022) found in a survey of 254 students at a historically Black university that by enabling an environment of connectiveness for students, the institution was better able to mitigate pandemic-related student mental stress.

What is the Secret Sauce?

The Mission of Black Identity Formation

Historically Black colleges and universities were created and run by the descendants of American Negro slaves. Historically Black colleges and universities have long been associated with broader goals than just conferring degrees on graduates, including a broader focus on individual and community prosperity. Their role in advocating for social justice and activism over time, going back decades to the period of legalized segregation, cannot be overestimated. It complements what students learn in the classroom, expanding their broader skills. This broader mission not only imparts on students a sense of belonging at their respective institutions but also within society.

While the mission of historically Black colleges and universities has not changed (Wright 2016), these institutions now serve a distinct student population—for the most part, Black students who *choose* to go to a historically Black institution despite having the option to attend a college or university that is not historically Black. As alluded to earlier in this essay (for example, Table 1), a significant proportion of these students increasingly come from disadvantaged backgrounds, and historically Black institutions continue to do well by them.

For all of these reasons, it is thus not surprising that graduates of these institutions seem more civically engaged and more likely to report that they are thriving in purpose and well-being (recall the polling results of Gallup [2015], to which we return below as well). The overrepresentation of graduates of historically Black colleges and universities in occupations that are plausibly positively correlated with confidence and self-esteem, such as US congressional representatives, court judges, university professors (as described in Fryer and Greenstone 2010; Toldson et al. 2022), and civil rights activists (as discussed in Redd 1998; Roebuck and Murty 1993), suggests that historically Black institutions have a comparative advantage in cultivating these traits.

Here, our focus is on how identity formation, confidence, and self-esteem boost academic achievement. Price, Spriggs, and Swinton (2011) provide a theoretical framework in which a college seeks to produce an ideal student identity through mechanisms such as curricular and cocurricular activities. The above-mentioned comparative advantage stems from the fact that historically Black colleges and universities have a higher ratio of Black to non-Black students, thus making it easier and less costly for Black students to subscribe to a particular identity and, in turn, to build high confidence, self-esteem, and self-image. This identity interacts with the choices that students make about investing energy and effort in their studies. This theoretical framework both rationalizes how historically Black institutions confer relative labor market advantages upon their graduates and constitutes a causal framework that enables historically Black colleges and universities to "punch above their weight" in general.

Gains in confidence and self-esteem can be viewed as the provision of social capital. Such traits, when successfully transmitted to students, can translate into individual productivity, including the observed patterns of academic success while in college and earnings in the labor market discussed earlier. Such noncognitive traits are of course difficult to measure, but a growing body of evidence for students and graduates of historically Black colleges and universities, some based on interviews, confirms their importance.

For example, "stereotype threat" refers to the common finding that if Black students are reminded of their racial identity before taking a test, they tend to score lower. However, when Alston et al. (2022) replicated the methods of these previous tests in a lab experiment at a historically Black university, they found no evidence of stereotype threat on test scores in this group. When McGee et al. (2021) interview 13 presidents or senior administrators at historically Black colleges and universities, they find that race-consciousness led to improved student performance by prioritizing the hiring of Black faculty, enhancing Black student experiences, and creating culturally affirming programming. Winkle-Wagner et al. (2020) discuss how alumnae credited Spelman College for fostering a culture of success.

In studies focused on graduate students, Tiako, Wages, and Perry (2022) find that Black medical students at historically Black institutions (1) report a greater sense of belonging than their Black colleagues at primarily white institutions, with a gap that widens over time, and (2) remain more confident in their scholastic capabilities (also see Gasman et al. 2017). Shorette and Palmer (2015) interview six Black men in graduate programs at a primarily white midwestern university and find evidence that those who graduated from historically Black institutions have an advantage in certain noncognitive traits including positive self-concept, realistic self-appraisal, ability to handle the systems they face, availability of a strong support person, leadership experience, and community participation. Similarly, Jett (2013) sought out four Black men who were graduate students in mathematics at various institutions. It turned out that all four had graduated from historically Black colleges and universities, and all of them emphasized the importance of supportive structures and mechanisms, especially the role models provided by African American male mathematics professors.

Looking at this body of evidence as a whole, both at the qualitative data and the overall record of success for both Black and other students, Smith and Jackson (2021) argue that historically Black colleges and universities can serve as a model for other institutions of higher education in the ways that they "(1) foster familial learning environments, (2) promote diversity and inclusion, and (3) invest more time, support, and compassion into students."

Table 3
Black Graduates: Support and Experiential Learning Opportunities

Outcome (Percent who strongly agree with the following statements)	*HBCU*	*Non-HBCU*
My professors at My University cared about me as a person	58%	25%
I had at least one professor at My University who made me excited about learning	74%	62%
While attending My University, I had a mentor who encouraged me to pursue my goals and dreams	42%	23%
Felt Support	35%	12%
While attending My University, I had an internship or job that allowed me to apply what I was learning in the classroom	41%	31%
While attending My University, I worked on a project that took a semester or more to complete	36%	30%
I was extremely active in extracurricular activities and organizations while attending My University	32%	23%
Experiential Learning	13%	7%

Source: Gallup (2015, p. 18).
Note: This table presents the percentage of Black graduates who strongly agree with a specific statement across historically Black colleges and universities and non-HBCUs.

Student Support from Faculty and Alumni

As reported by Gallup (2015, p. 18) and shown in Table 3, Black graduates of historically Black colleges and universities recall more positive and supportive experiences than Black graduates of other institutions. Black graduates of historically Black institutions are more than twice as likely to have felt supported. Digging into this data more deeply, they are also more likely to recall involvement in applied internships, long-term projects, and extracurricular activities. The gap between Black graduates of historically Black institutions and Black graduates of other colleges and universities is 58 percent versus 25 percent in recalling having professors who cared about them as a person, and 42 percent to 23 percent in recalling that they had a mentor who encouraged them to pursue their goals and dreams. Similarly, the Baccalaureate and Beyond data suggest that Black graduates of historically Black institutions are more likely to have participated in cocurricular experiences, such as a paid internship.

The role of alumni in the success of historically Black institutions should not be underestimated. This support ranges from donations and mentoring (both in and out of the classroom) to exerting pressure on college and university leadership to stay true to their mission. In fact, many alumni give back by returning to their institutions, particularly as faculty. In other instances, alumni are parents, grandparents, aunts, uncles, and so on of current students at historically Black institutions. Based on the 2016–2017 Baccalaureate and Beyond cohort, 11.2 percent of Black students at historically Black institutions utilized the alumni network during their undergraduate education, relative to 8.5 percent of Black students at other colleges and universities. These statistics are probably underestimates, because alumni who return during job market season to recruit students are often not reported as part of "utilizing the alumni network." Also, networks of parents and relatives are unlikely to be reported, even when they are alumni as well.

The above findings are consistent with other related work. For example, Palmer, Davis, and Maramba (2010) investigate the academic and social experiences of eleven Black males, who entered a public historically Black institution through its remedial or developmental studies program and persisted to graduation. Participants credited the university's racial composition, support from peers, faculty, and role models in helping to increase their propensity for learning and academic success. Moreover, there is a growing literature that underscores the importance of Black teachers being trained at historically Black institutions, since such training seems to differentially impact Black student achievement (for example, Edmonds 2022; Morgan and Hu 2022; Redding 2019).

In the rest of this section, we highlight some of the ways in which historically Black colleges and universities have managed to pave the way in curricular and cocurricular innovation for Black students: college preparation, Black-centered curriculum and first-year courses, pathways to college, research experiences, and pathways to terminal degrees. Our discussion is only meant to offer some concrete illustrations, not to be exhaustive. The list covered here primarily reflects specific initiatives with which we are most familiar.

College Preparation

Many historically Black colleges and universities have college preparation summer programs that seek to expose local youth to their mission, programs, campus—and to Black culture more generally. HBCU Buzz (https://hbcubuzz.com/category/opportunities/internships/, accessed on April 8, 2023) and HBCU First (https://hbcufirst.com/, accessed on April 8, 2023) curate lists of select summer programs, while institutional websites provide up-to-date information for specific programs. A specific example is Morehouse College's Summer Academy (https://bit.ly/3bl708Q, accessed on April 8, 2023).

Summer programs benefit both current and prospective students at historically Black institutions. Current students are often involved with such programs; for example, as teaching, research, and/or resident assistants. This involvement helps current students develop technical and soft skills that serve them well on the job market and when applying for graduate school. It also helps current students with stronger letters of recommendation and references.

For prospective students—and Black youth in particular—such programs can be a series of firsts: for example, first time setting foot on a college campus for a period longer than a few hours and first time encountering faculty (teachers), staff, and peers who mostly look like them. With these experiences comes the beginning of an understanding of the culture of historically Black institutions: rigorous academics in a supportive community environment, and how that combines with a social life on and off campus.

Black-Centered Curricula and First-Year Experiences

Historically Black colleges and universities tend to tailor their curricula, particularly during the first year, to Black culture. Such an Afrocentric approach has been found to be complementary to effective learning outcomes for Black students (for example, as described in the literature review by Gasman and Commodore 2014; or in the "living-learning culture" discussed by Ericksen and Williamson-Ashe 2019).

An example of a Black-centered curriculum is the "African Diaspora and the World" (ADW) course at Spelman College, a two-semester sequence required in the first year or for transfer students. The sequence speaks to students' experiences as Black women and in so doing, helps them learn about themselves, their history, and their place in the African diaspora and the world. Many alumnae refer to the course as being the most formative educational influence in their lives (per the course website, https://bit.ly/3N8opyY, accessed on April 8, 2023). ADW has existed for more than 25 years and tends to challenge students' critical inquiry and reading skills. From 2015 to 2020, a multi-disciplinary team of Spelman faculty and staff, supported by a grant from the US Department of Education, implemented a randomized controlled trial in collaboration with the ADW course. The main purpose of the intervention was to test whether metacognitive strategies could further enhance student learning in challenging first-year courses and have follow-on impacts as students progress through the college curriculum. Faculty were randomized to either a treatment condition in which they were trained to teach the material using metacognitive strategies such as reciprocal teaching, or a control condition in which they taught ADW as they usually would have in absence of the intervention (for an overview of the research, see Blankson et al. 2019). This experiment is a testament to historically Black institutions' commitment to the scholarship of teaching and learning as well as providing Black students with the best curricular and cocurricular experience.

A concrete example of a first-year experience program is the one at Winston-Salem State University. It is designed to support first-year students in their transition from high school to college, and provide them with a comprehensive, holistic experience that creates successful learners and competent citizens. The experience is centered around three keywords: empower, engage, and explore. First-year students are empowered through enhanced knowledge about communication, time and money management, effective study strategies, and research. They engage with peers, faculty, staff, and administrators through special activities, volunteering, and activism. Finally, they explore social justice in five areas: diversity, educational equity, health equity, community empowerment, and community sustainability. Additional information about the Winston-Salem first-year experience course is available at https://bit.ly/3Ng9IKh (accessed on April 8, 2023).

Research Experiences in and out of the Classroom

Many historically Black colleges and universities have long embraced research experiences, both in and out of the classroom. We discuss a few examples to illustrate, although we believe that they typify the experiences at many other institutions.

First, the Course-based Undergraduate Research Experiences (CURE) project at Xavier University of Louisiana is supported by a Targeted Infusion Project award from the National Science Foundation's (NSF) HBCU program. Per the NSF awards page (https://bit.ly/3zTHRwj, accessed on April 8, 2023), this CURE project (expected to end in 2023) infuses two authentic research projects in first-year biology labs. The key objectives of the CURE are to (1) increase students' scientific competencies, motivation, and self-efficacy and (2) broaden their knowledge of career options. Data on student learning, self-efficacy, project-ownership, and retention will be analyzed to further advance science education at historically Black institutions.[3]

The Research Initiative for Scientific Enhancement (RISE) program at Morgan State University seeks to equip scholars with resilience, excellence, achievement, and community in health sciences by adopting a holistic approach to college matriculation and graduate training preparation. The RISE program provides a supportive community that features hands-on research training, mentoring, academic support, and science identity development. These strategies are intended to strengthen student preparedness to enter PhD training in the biomedical sciences with the long-term goal of contributing researchers that possess the ability to address some of the nation's critical health disparities (per the program website, https://bit.ly/3br2cPG, accessed on April 8, 2023). A related set of programs are funded by the National Institutes of Health, https://bit.ly/3bnjEnR (accessed on April 8, 2023).

Some historically Black institutions, perhaps especially Morehouse and Spelman Colleges, have been successful at utilizing external (sometimes federal) funding to enable student research experiences that make peer-reviewed published contributions to knowledge. In our own experience, Smith and Viceisza (2018) includes one undergraduate student as co-author, and was supported by the Kauffman Foundation (also see Viceisza's lab at Spelman, VLab, http://bit.ly/2CTseGs, accessed on April 2, 2023, which involves students in research experiences more generally). Elu et al. (2019) includes nine undergraduate students as co-authors and was supported by NSF award 1748433 (https://bit.ly/3boNGI2, accessed on April 8, 2023).

Pathways to Terminal Degrees

As mentioned previously, almost one-quarter of Black graduates who earned a doctorate in science and engineering between 2015 and 2019 earned their bachelor's degree from a historically Black college or university (National Center for Science and Engineering Statistics 2021).

Here, we discuss two prominent examples of pathways to graduate school and terminal degrees. The Fisk-Vanderbilt Master's-to-PhD Bridge Program exists to increase the number of minority students engaged in PhD-level research in science

[3] Lee and Searles (2021) discuss the IBM-HBCU Quantum Center, which is a collaboration between IBM and a consortium of 23 historically Black institutions that seeks to address the lack of Black representation and build a diverse and aware workforce in quantum information science and engineering. One key pillar of the Center is to provide funding to support undergraduate, graduate, and faculty research.

(per their website, https://bit.ly/3OLGCnn, accessed on April 8, 2023). All students begin by working toward a master's degree in physics, biology, or chemistry under the guidance of faculty mentors, so that they can develop the academic foundation, research skills, and one-on-one mentoring relationships that will foster a successful transition to the PhD. According to Stassun, Burger, and Lange (2010), the program couples targeted recruitment with active retention strategies, and is built upon a clearly defined structure that is flexible enough to address individual student needs while maintaining clearly communicated baseline standards for student performance. A key precept of the program's philosophy is to eliminate passivity in student mentoring; students are deliberately groomed to successfully transition into the PhD program through active involvement in research experiences with future PhD advisers, coursework that demonstrates competency in core PhD subject areas, and frequent interactions with joint mentoring committees. While geographical proximity of the implementing institutions seems important, it may not be insurmountable given the possibility of virtual interactions. To date, 152 students have enrolled in the program, 118 master's degrees have been awarded, 100 students have bridged to PhD programs, and 40 students have earned the PhD, 30 of those from Vanderbilt. Today, 31 students are in a Vanderbilt PhD program, while 17 are in a Fisk Master's program. Fifty-one percent of the students are African American, 22 percent are Hispanic, 7 percent are other minority, including Native Hawaiian and Pacific Islander, and 20 percent are white or other nonminority. Fifty-seven percent are female.

Prairie View's Department of Biology has created two successful programs aimed at producing Black doctors (Gasman et al. 2017). The Premedical Concepts Institute is a rigorous ten-week summer program for incoming first-year students interested in pursuing science careers. The Cardiovascular and Microbial Research Center provides undergraduate students with research projects and mentoring that support independent problem solving. One of the institution's approaches to ensuring that students see themselves as successful is to bring former students—now successful alumni serving as doctors across the nation—back to campus at the beginning of the academic year and throughout students' academic programs. This strategy has been echoed at many historically Black colleges and universities and underscores at least a couple of issues highlighted previously—the importance of alumni and role models.

Where Do We Go from Here?

Since their founding, dating back in many cases to the nineteenth century, historically Black colleges and universities have played a significant role in the education and social mobility of African Americans. They have lessons to teach about how to educate Black college students, as well as a continuing role to play as part of the ecosystem of US higher education.

About 90 percent of Black college students do not attend a historically Black institution. In addition, if the number of Black students attending college is to rise substantially in the future, many of the additions will come on the margin of high school students with lower socioeconomic status and less academic preparation—the margin where historically Black institutions have been having success. Our discussion suggests a number of ways in which institutions that are not traditionally Black can encourage and support Black students: (1) introduce high school and post-college bridge programs with a focus on inclusion of Black students; (2) tailor parts of their curriculum and first-year experiences in particular to Black culture; (3) engage Black students with cocurricular experiences on research, social justice, and activism; and (4) expose Black students to role models who look like them, including Black alumni and faculty (the latter would of course require these other institutions to hire a critical mass of such faculty to begin with, an issue that reaches beyond the scope of this essay).

In addition, and perhaps more importantly, we believe that there are fruitful opportunities for collaboration between historically Black and predominantly white colleges and universities, like the Fisk-Vanderbilt Master's-to-PhD Program discussed earlier. In addition, we believe in the possibilities for faculty exchange programs that go in both directions. Some student and faculty exchanges exist now, but it is our impression that many more could do so.

The reason to emphasize these types of collaborations is that some ingredients to the secret sauce of historically Black colleges and universities are organic to the institutions themselves: they were founded as institutions intended to educate Black Americans and in so doing have a comparative advantage that enables them to impart on Black students a sense of belonging, increased confidence and self-esteem, and skills that benefit the Black community and society at large. They have built expertise at reaching out to the disadvantaged, and their initiatives to do so have developed organically—in alignment with the institution's mission—over time. It may be difficult, if not impossible, to replicate such efforts in primarily white institutions. Ad hoc attempts at "diversity, equity, and inclusion" efforts can feel forced, giving rise to concerns of tokenism and other types of stigma, stereotype threat, and pandering.

Historically Black colleges and universities continue to play an important role in the broader ecosystem of US higher education. Their continuing vitality might be boosted along two dimensions: managerial and financial. The quality of executive leadership, particularly the president of a college or university, may matter for student outcomes (for example, Kuh et al. 2005; Friedman and Kass-Shraibman 2017). However, executive leadership at historically Black institutions seems to be disproportionately challenged and contested. For example, in the data collected by Sean McKinniss on no-confidence votes for college presidents from 1989 to 2021 (https://bit.ly/3A4LnnH, accessed on April 8, 2023), historically Black institutions account for approximately 6 percent of no-confidence votes while constituting only 3 percent, approximately, of all US colleges and universities.

Finally, a substantial boost in funding for historically Black colleges and universities seems appropriate and overdue. These institutions were dramatically underfunded on a per-student basis during decades of legalized segregation (Smith 2021). Colleges and universities are long-lived institutions, and underfunding from decades ago will affect the physical inheritance of these institutions today: academic buildings, library resources, greenspaces, athletic facilities, and more. This underfunding has continued up to the present. Lawsuits about underfunding of historically Black institutions have been settled in recent years by Maryland, Alabama, and Mississippi, and a similar lawsuit was recently announced in Florida. Two journalists at *Forbes* calculated that Black land-grant universities have been underfunded at the state level, relative to their primarily white counterparts in the same states, by $12.8 billion in the last three decades (Adams and Tucker 2022).

In part due to the killing of George Floyd in May 2020, there have recently been some notable private and public investments in historically Black colleges and universities. MacKenzie Scott has given $560 million in the last few years (as reported by Freeman 2021) and the American Rescue Plan passed in 2022 provides $2.7 billion for historically Black institutions (White House 2022). Still, this is only a drop in the bucket given decades of underfunding to historically Black institutions (Smith 2021) and centuries of marginalization of the broader African-American community (Baker 2022). Black colleges and universities generate a good return on investment (for additional discussion, see Morse, Sakano, and Price 1996; Sharpe 2016). If the United States aspires to improve the education, social mobility, and future of African Americans, it should act to close the financial resource gap that has long prevailed between historically Black institutions and other colleges and universities.

■ *We are thankful to Jason Owen-Smith, Jonathan Smith, Omari Swinton, Bruce Weinberg, and the editors for helpful discussions. Viceisza is grateful to the Hoover Institution at Stanford University where he was a W. Glenn Campbell and Rita Ricardo-Campbell National Fellow and the John Stauffer National Fellow when this project was initiated. In 2021–2022, Viceisza was also a Carnegie Corporation and Rockefeller Foundation Distinguished Researcher and Creative Scholar at Spelman College.*

References

Adams, Susan, and Hank Tucker. 2022. "How America Cheated Its Black Colleges." *Forbes*. https://www.forbes.com/sites/susanadams/2022/02/01/for-hbcus-cheated-out-ofbillions-bomb-threats-are-latest-indignity/?sh=2ce6fd6a640c (accessed April 8, 2023).

Agesa, Jacqueline, Maury Granger, and Gregory N. Price. 2000. "Economics Faculty Research at Teaching Institutions: Are Historically Black Colleges Different?" *Southern Economic Journal* 67 (2): 427–47.

Akerlof, George A., and Rachel E. Kranton. 2002. "Identity and Schooling: Some Lessons for the Economics of Education." *Journal of Economic Literature* 40 (4): 1167–1201.

Alston, Mackenzie, William A. Darity, Catherine C. Eckel, Lawrence McNeil, and Rhonda Sharpe. 2022. "The Effect of Stereotypes on Black College Test Scores at a Historically Black University." *Journal of Economic Behavior and Organization* 194: 408–24.

Baker, Regina S. 2022. "The Historical Racial Regime and Racial Inequality in Poverty in the American South." *American Journal of Sociology* 127 (6): 1721–81.

Biasi, Barbara, and Song Ma. 2022. "The Education-Innovation Gap." NBER Working Paper 29853.

Blankson, A. Nayena, Jimmeka J. Guillory, Francesina R. Jackson, Angelino C. G. Viceisza, and Bruce H. Wade. 2019. "Lessons Learned from a Multidisciplinary Randomized Controlled Trial." *Review of Black Political Economy* 46 (2): 160–66.

Braby, Laura, Matthew Holcomb, and Christoph Leonhard. 2022. "Examining the Relationship between Ethnic Identity, Depression, and Alcohol Use among Students at Historically Black Colleges/Universities (HBCUs)." *Journal of Ethnicity in Substance Abuse* 21 (2): 621–37.

Broady, Kristen E., Curtis L. Todd, and Darlene Booth-Bell. 2017. "Dreaming and Doing at Georgia HBCUs: Continued Relevancy in 'Post-racial' America." *Review of Black Political Economy* 44 (1–2): 37–54.

Colen, Cynthia G., Nicolo P. Pinchak, and Kierra S. Barnett. 2021. "Racial Disparities in Health among College-Educated African Americans: Can Attendance at Historically Black Colleges or Universities Reduce the Risk of Metabolic Syndrome in Midlife?" *American Journal of Epidemiology* 190 (4): 553–61.

Constantine, Jill M. 1995. "The Effect of Attending Historically Black Colleges and Universities on Future Wages of Black Students." *ILR Review* 48 (3): 531–46.

de Zeeuw, Mels, Sameera Fazili, and Julie L. Hotchkiss. 2021. "Decomposing Differences in Black Student Graduation Rates between HBCU and Non-HBCU Institutions: The Devil Is in the Details." *Economics Letters* 202: 109816.

Edmonds, Lavar. 2022. "Role Models Revisited: HBCUs, Same-Race Teacher Effects, and Black Student Achievement." Unpublished.

Elu, Juliet U., Jared Ireland, David Jeffries, Ivory Johnson, Ellis Jones, Dimone Long, Gregory N. Price et al. 2019. "The Earnings and Income Mobility Consequences of Attending a Historically Black College/University: Matching Estimates from 2015 U.S. Department of Education College Scorecard Data." *Review of Black Political Economy* 46 (3): 171–92.

Ericksen, Kirsten S., and Sandra Williamson-Ashe. 2019. "Holistic Justice Instruction through High-Impact Educational Practices at Historically Black Colleges and Universities." *Urban Social Work* 3 (S1): S100–114.

Francis, Dania V., Bradley L. Hardy, and Damon Jones. 2022. "Black Economists on Race and Policy: Contributions to Education, Poverty and Mobility, and Public Finance." *Journal of Economic Literature* 60 (2): 454–93.

Freeman, Tyrone McKinley. 2021. "MacKenzie Scott's HBCU Giving Starkly Contrasts with the Approach of Early White Funders of Historically Black Colleges and Universities." *The Conversation*, August 2. https://theconversation.com/mackenzie-scotts-hbcu-giving-starkly-contrasts-with-the-approach-of-early-white-funders-of-historically-black-colleges-and-universities-159039.

Friedman, Hershey H., and Frimette Kass-Shraibman 2017. "How to Pad Your CV. Inside Higher Ed." http://bit.ly/3UnXq7U. Accessed on April 8, 2023.

Fryer, Roland G., Jr., and Michael Greenstone. 2010. "The Changing Consequences of Attending Historically Black Colleges and Universities." *American Economic Journal: Applied Economics* 2 (1): 116–48.

Gallup. 2015. *Gallup-USA Funds Minority College Graduates Report*. Washington, DC: Gallup, Inc.

Gasman, Marybeth, and Felecia Commodore. 2014. "The State of Research on Historically Black Colleges and Universities." *Journal for Multicultural Education* 8 (2): 89–111.

Gasman, Marybeth, Tiffany Smith, Carmen Ye, and Thai-Huy Nguyen. 2017. "HBCUs and the Production of Doctors." *AIMS Public Health* 4 (6): 579–89.

Gordon, Ethan K., Zackary B. Hawley, Ryan Carrasco Kobler, and Jonathan C. Rork. 2021. "The Paradox of HBCU Graduation Rates." *Research in Higher Education* 62 (3): 332–58.

Hammond, Miriam, LaToya Owens, and Brian Gulko. 2021. *HBCUs Transforming Generations: Social Mobility Outcomes for HBCU Alumni*. Washington, DC: UNCF.

Hill, Nicholas J., Haydar Kurban, William Spriggs, and Omari H. Swinton. Forthcoming. "An Empirical Analysis of HBCU Attendance on Black Student's Graduation Rates." *Journal of Negro Education*.

Huang, Hsuan Yuan, Huijun Li, and Ying-Chia Hsu. 2022. "Coping, COVID Knowledge, Communication, and HBCU Student's Emotional Well-Being: Mediating Role of Perceived Control and Social

Connectedness." *Journal of Community Psychology* 50 (6): 2703–25.

Itzkowitz, Michael. 2022. "Out with the Old, In with the New: Rating Higher Ed by Economic Mobility." Washington DC: Third Way.

Jett, Christopher C. 2013. "HBCUs Propel African American Male Mathematics Majors." *Journal of African American Studies* 17 (2): 189–205.

Kuh, George D., Jillian Kinzie, John H. Schuh, and Elizabeth J. Whitt. 2005. "Never Let It Rest Lessons about Student Success from High-Performing Colleges and Universities." *Change: The Magazine of Higher Learning* 37 (4): 44–51.

Lee, Kayla B., and Thomas A. Searles. 2021. "IBM-HBCU Quantum Center: A Model for Industry-Academic Partnerships to Advance the Creation of a Diverse, Quantum Aware Workforce." In *2021 IEEE International Conference on Quantum Computing and Engineering* (*QCE*), ed. Hausi A. Müller, Greg Byrd, Candace Culhane, and Travis Humble, 392–396. Los Alamitos, CA: IEEE Computer Society.

McGee, Ebony O., Lynette Parker, Orlando L. Taylor, Kelly Mack, and Margaret Kanipes. 2021. "HBCU Presidents and Their Racially Conscious Approaches to Diversifying STEM." *Journal of Negro Education* 90 (3): 288–305.

Morgan, Paul L., and Eric Hengyu Hu. 2022. "Fixed Effect Estimates of Student-Teacher Racial or Ethnic Matching in U.S. Elementary Schools." *Early Childhood Research Quarterly* 63: 98–112.

Morse, Lawrence, Ryoichi Sakano, and Gregory N. Price. 1996. "Black Public Colleges and Universities as Projects: How Do They Rank Relative to White Public Colleges and Universities?" *Review of Black Political Economy* 24 (4): 65–79.

Nathenson, Robert A., Andrés Castro Samayoa, and Marybeth Gasman. 2019. *Moving Upward and Onward: Income Mobility at Historically Black Colleges and Universities.* New Brunswick, NJ: Rutgers Center for Minority Serving Institutions.

National Center for Education Statistics. 2022. "Digest of Education Statistics 2022." National Center for Education Statistics. https://nces.ed.gov/programs/digest/2022menu_tables.asp.

National Center for Education Statistics. 2023. "Fast Facts: Historically Black Colleges and Universities." National Center for Education Statistics. https://nces.ed.gov/fastfacts/display.asp?id=667 (accessed February 10, 2023).

National Center for Science and Engineering Statistics. 2021. *Women, Minorities, and Persons with Disabilities in Science and Engineering.* NSF 21-321. Alexandria, VA: National Science Foundation.

Palmer, Robert and Marybeth Gasman. 2008. "'It Takes a Village to Raise a Child': The Role of Social Capital in Promoting Academic Success for African American Men at a Black College." *Journal of College Student Development* 49 (1): 52–70.

Palmer, Robert T., Ryan J. Davis, and Dina C. Maramba. 2010. "Role of an HBCU in Supporting Academic Success for Underprepared Black Males." *Negro Educational Review* 61 (1–4): 85–106.

Porter, Catherine and Danila Serra. 2020. "Gender Differences in the Choice of Major: The Importance of Female Role Models." *American Economic Journal: Applied Economics* 12 (3): 226–54.

Presidential Committee on the Legacy of Slavery. 2022. *The Legacy of Slavery at Harvard: Report and Recommendations of the Presidential Committee.* Cambridge, MA: Harvard University Press.

Price, Gregory N., and Angelino C. G. Viceisza. 2023. "Replication data for: What Can Historically Black Colleges and Univesities Teach about Improving Higher Education Outcomes for Black Students?" American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E188461V1.

Price, Gregory N., and Rhonda V. Sharpe. 2020. "Is the Economics Knowledge Production Function Constrained by Race in the USA?" *Journal of the Knowledge Economy* 11 (2): 614–29.

Price, Gregory N., William Spriggs, and Omari H. Swinton. 2011. "The Relative Returns to Graduating from a Historically Black College/University: Propensity Score Matching Estimates from the National Survey of Black Americans." *Review of Black Political Economy* 38 (2): 103–30.

Redd, Kenneth E. 1998. "Historically Black Colleges and Universities: Making a Comeback." *New Directions for Higher Education* 102: 33–43.

Redding, Christopher. 2019. "A Teacher Like Me: A Review of the Effect of Student–Teacher Racial/Ethnic Matching on Teacher Perceptions of Students and Student Academic and Behavioral Outcomes." *Review of Educational Research* 89 (4): 499–535.

Richards, David A. R., and Janet T. Awokoya. 2010. *Understanding HBCU Retention and Completion.* Washington, DC: UNCF Frederick D. Patterson Research Institute.

Roebuck, Julian B., and Komanduri S. Murty. 1993. *Historically Black Colleges and Universities: Their Place* *in American Higher Education*. Westport, CT: Praeger.

Saunders, Katherine M., and Barry T. Nagle. 2019. *HBCUs Punching above Their Weight: A State-Level Analysis of Historically Black College and University Enrollment Graduation*. Washington, DC: UNCF Frederick D. Patterson Research Institute.

Sharpe, Rhonda Vonshay. 2016. "HBCUs: Efficiencies of Creating a Scientific Workforce Outta Fifteen Cents." In *Administrative Challenges and Organizational Leadership in Historically Black Colleges and Universities*, edited by Charles B. W. Prince and Rochelle L. Ford, 199–212. Hershey: IGI Global.

Shorette, C. Rob, II, and Robert T. Palmer. 2015. "Historically Black Colleges and Universities (HBCUs): Critical Facilitators of Non-cognitive Skills for Black Males." *Western Journal of Black Studies* 39 (1): 18–29.

Simkins, Scott, and Stuart Allen. 2000. "Are Learning Outcomes in Economics Different at Predominantly Black and White Universities? Lessons from Principles of Macroeconomics Courses at Two Schools." *Review of Black Political Economy* 28 (3): 23–39.

Smith, Baylee, and Angelino Viceisza. 2018. "Bite Me! ABC's Shark Tank as a Path to Entrepreneurship." *Small Business Economics* 50 (3): 463–79.

Smith, Denise A. 2021. *Achieving Financial Equity and Justice for HBCUs.* New York: Century Foundation.

Smith, Jennifer M., and Elliot O. Jackson. 2021. "Historically Black Colleges & Universities: A Model for American Education." *Florida A&M University Law Review* 14 (1): 103–239.

Stassun, Keivan Guadalupe, Arnold Burger, and Sheila Edwards Lange. 2010. "The Fisk-Vanderbilt Masters-to-PhD Bridge Program: A Model for Broadening Participation of Underrepresented Groups in the Physical Sciences through Effective Partnerships with Minority-Serving Institutions." *Journal of Geoscience Education* 58 (3): 135–44.

Tiako, Max Jordan Nguemeni, James E. Wages, and Sylvia P. Perry. 2022. "Black Medical Students' Sense of Belonging and Confidence in Scholastic Abilities at Historically Black vs Predominantly White Medical Schools: A Prospective Study." *Journal of General Internal Medicine* 38 (1): 122–24.

Toldson, Ivory A., Mercy Mugo, Jennifer Hudson, Mahlet Megra, and Cynthia Overton. 2022. "HBCUs as a Pathway to Becoming a Scientist: Institutional Characteristics of HBCUs That Are among the Top Baccalaureate Origins of Black Doctorate Recipients in STEM." In *Black College Leadership in PK-12 Education*, edited by Ivory A. Toldson, 98–122. Leiden: Koninklijke Brill.

White House. 2022. "Fact Sheet: State-by-State Analysis of Record $2.7 Billion American Rescue Plan Investment in Historically Black Colleges and Universities." March 7. https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/07/fact-sheet-state-by-state-analysis-of-record-2-7-billion-american-rescue-plan-investment-in-historically-black-colleges-and-universities/.

Wilson, Valerie Rawlston. 2007. "The Effect of Attending an HBCU on Persistence and Graduation Outcomes of African-American College Students." *Review of Black Political Economy* 34 (1–2): 11–52.

Winkle-Wagner, Rachelle, Jacqueline M. Forbes, Shelby Rogers, and Tangela Blakely Reavis. 2020. "A Culture of Success: Black Alumnae Discussions of the Assets-Based Approach at Spelman College." *Journal of Higher Education* 91 (5): 653–73.

Wood, J. Luke, and Robert T. Palmer. 2017. "HBCU Labor Market Outcomes: An Examination of Baccalaureate Degree Holders' Earnings and Benefits." In *Black Colleges across the Diaspora: Global Perspectives on Race and Stratification in Postsecondary Education*, Vol. 14, *Advances in Education in Diverse Communities*, edited by M. Christopher Brown II and T. Elon Dancy II, 171–86. Bingley, UK: Emerald Publishing.

Wright, Mary. 2016. "Mission Accomplished: The Unfinished Relationship between Black Law Schools and Their Historical Constituencies." *North Carolina Central Law Review* 39 (1): Article 3, 1–32.

Journal of Economic Perspectives—Volume 37, Number 3—Summer 2023—Pages 233–240

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by e-mail at <taylort@macalester.edu>, or c/o Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., Saint Paul, MN 55105.*

Smorgasbord

The Federal Reserve has published a "Review of the Federal Reserve's Supervision and Regulation of Silicon Valley Bank" (April 2023, https://www.federalreserve.gov/publications/review-of-the-federal-reserves-supervision-and-regulation-of-silicon-valley-bank.htm), focused on events leading up to the closure of the Silicon Valley Bank Financial Group on March 9, 2023. "Uninsured depositors interpreted SVBFG's announcements on March 8 as a signal that [the firm] was in financial distress and began withdrawing deposits on March 9, when SVB experienced a total deposit outflow of over $40 billion. This run on deposits at SVB appears to have been fueled by social media and SVB's concentrated network of venture capital investors and technology firms that withdrew their deposits in a coordinated manner

■ *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.37.3.233.

with unprecedented speed. On the evening of March 9 and into the morning of March 10, SVB communicated to supervisors that the firm expected an additional over $100 billion in outflows during the day on March 10. SVB did not have enough cash or collateral to meet the extraordinary and rapid outflows. . . . This deposit outflow was remarkable in terms of scale and scope and represented roughly 85 percent of the bank's deposit base. By comparison, estimates suggest that the failure of Wachovia in 2008 included about $10 billion in outflows over 8 days, while the failure of Washington Mutual in 2008 included $19 billion over 16 days. . . . The full board of directors did not receive adequate information from management about risks at Silicon Valley Bank and did not hold management accountable for effectively managing the firm's risks. The bank failed its own internal liquidity stress tests and did not have workable plans to access liquidity in times of stress. Silicon Valley Bank managed interest rate risks with a focus on short-run profits and protection from potential rate decreases, and removed interest rate hedges, rather than managing long-run risks and the risk of rising rates."

Stephen J. Ceci, Shulamit Kahn, and Wendy M. Williams have written "Exploring Gender Bias in Six Key Domains of Academic Science: An Adversarial Collaboration" (*Psychological Science in the Public Interest*, online April 26, 2023, https://journals.sagepub.com/doi/full/10.1177/15291006231163179). "This article represents more than 4.5 years of effort by its three authors. By the time readers finish it, some may assume that the authors were in agreement about the nature and prevalence of gender bias from the start. However, this is definitely not the case. Rather, we are collegial adversaries who, during the 4.5 years that we worked on this article, continually challenged each other, modified or deleted text that we disagreed with, and often pushed the article in different directions. Although the three of us have exchanged hundreds of emails and participated in many Zoom sessions, Kahn has never met Ceci and Williams in person. . . . We synthesized the vast, contradictory scholarly literature on gender bias in academic science from 2000 to 2020. . . . We evaluated the empirical evidence for gender bias in six key contexts in the tenure-track academy: (a) tenure-track hiring, (b) grant funding, (c) teaching ratings, (d) journal acceptances, (e) salaries, and (f) recommendation letters. We also explored the gender gap in a seventh area, journal productivity, because it can moderate bias in other contexts. We focused on these specific domains, in which sexism has most often been alleged to be pervasive, because they represent important types of evaluation, and the extensive research corpus within these domains provides sufficient quantitative data for comprehensive analysis. Contrary to the omnipresent claims of sexism in these domains appearing in top journals and the media, our findings show that tenure-track women are at parity with tenure-track men in three domains (grant funding, journal acceptances, and recommendation letters) and are advantaged over men in a fourth domain (hiring). For teaching ratings and salaries, we found evidence of bias against women; although gender gaps in salary were much smaller than often claimed, they were nevertheless concerning."

Shahid Yusuf describes Korea's development experience in "Could Innovation and Productivity Drive Growth in African Countries? Lessons from Korea"

(Center for Global Development, Working Paper 635, March 2023, https://www.cgdev.org/publication/could-innovation-and-productivity-drive-growth-african-countries-lessons-korea). "Korea's swift ascent up the income ladder has at least three (overlapping) explanations. The unwavering commitment of the political leadership and the business elite, starting with President Park Chung Hee in the mid 1960s and sustained by his successors, to a relatively inclusive, export-led industrial strategy entailing systematic diversification into more complex manufactures, is arguably the most frequently retailed. The strategy itself was choreographed and implemented by Korea's economic bureaucracy headed by the Economic Planning Board (EPB) in consultation with the leading business groups. . . . The more 'neoclassical' explanation for why Korea pulled ahead of comparators in Asia and Africa sidesteps industrial policy and instead emphasizes the crafting of an enabling environment incentivizing and steering private investment into promising industries. This was complemented by public investment in energy and transport infrastructure and measures that deepened the skills of the workforce. This line of reasoning privileges market forces with the state playing an important supporting role, gives due attention to the initiative of private business conglomerates (chaebol) that spared no effort in penetrating foreign markets, and to (East Asian) neighborhood effects that conferred reputational advantages and attracted the attention of foreign buyers and investors. However, neither industrial policy nor market forces would have delivered the results Korea did achieve absent the great strides Korea made in absorbing and mastering technology from abroad and mustering home grown ST&I [science, technology, and innovation] capabilities. This third explanation intersects with and underpins the other reasons put forward. Unlike many other developing nations, Korea perceived and grasped technological opportunities and put them to good use. The state took the lead in creating the foundations of what was to become Korea's innovation system."

Steven A. Altman and Caroline R. Bastian have produced the *DHL Global Connectedness Index 2022*, subtitled "An in-depth report on globalization" (https://www.dhl.com/global-en/delivered/globalization/global-connectedness-index.html). Here are some of the key takeaways: "International flows have proven remarkably resilient through recent crises, strongly rebutting the notion that globalization has gone into reverse. . . . There is evidence of decoupling between the United States and China across most types of international flows. This decoupling has not—or at least not yet—led to a broader fragmentation of international activity between rival blocs. . . . Trade flows stretched out over longer distances during the Covid-19 pandemic . . . Roughly half of all international flows already happen inside major world regions, and it is still an open question whether regionalization will increase significantly in the coming years. . . . The volume of world trade in goods reached 10% above its pre-pandemic level in mid-2022, and trade in services also surpassed pre-pandemic levels last year. . . . Foreign direct investment flows, which reflect companies buying, building, or reinvesting in international operations, rebounded to above pre-pandemic levels in 2021, before starting to weaken in the second quarter of 2022. . . . The number of people traveling to foreign countries roughly doubled in 2022, but was still down 37% from 2019."

Fabian Villalobos, Jonathan L. Brosmer, Richard Silberglitt, Justin M. Lee, and Aimee E. Curtright make their case in "Time for Resilient Critical Material Supply Chain Policies" (Rand Corporation 2022, https://www.rand.org/pubs/research_reports/RRA2102-1.html). "China is the largest producer and processor of rare earth oxides (REOs) worldwide and a key producer of lithium-ion battery (LIB) materials and components. China's market share of REO extraction has decreased, but it still has large influence over the downstream supply chain–processing and magnet manufacturing. Chinese market share of the LIB supply chain mirrors REO supply bottlenecks. If it desired, China could effectively cut off 40 to 50 percent of global REO supply, affecting U.S. manufacturers and suppliers of DoD [Department of Defense] systems and platforms. Although a deliberate disruption is unlikely, resilience against supply disruption and building domestic competitiveness are important."

Antitrust

Timothy J. Muris, who was head of the Federal Trade Commission from 2001 to 2004, discusses "Neo-Brandeisian Antitrust: Repeating History's Mistakes" (AEI Economics Working Paper 2023-02, January 2023, https://www.aei.org/research-products/working-paper/neo-brandeisian-antitrust-repeating-historys-mistakes/). "In July 2021, with his top White House competition adviser and the new chair of the Federal Trade Commission (FTC) at his side as he introduced a competition executive order, President Joe Biden decried the 'experiment failed' in economics-driven antitrust over the past 40 years. The president promised to return to antitrust 'traditions' that existed before the 'failed' 40 years. The Biden appointees . . . call themselves neo-Brandeisians . . . Louis Brandeis's famous 1914 Harper's Weekly article, 'A Curse of Bigness,' inspired the title . . . To the neo-Brandeisians, big is again bad. Bad, not because it harms consumers—we will see that . . . harm to consumers is not the appropriate test for judging business conduct—but bad in some overarching political sense and for its own sake. . . . The chapters that follow this introduction study in detail two examples of past antitrust policy that the Biden antitrust leaders praise: vigorous enforcement of the Robinson-Patman Act and 1960s-style merger enforcement. . . . This, then, is the world of populist merger enforcement, the world that the neo-Brandeisians praise. It is a world in which the consumer-welfare standard does not exist. Consumers are so irrelevant that merging companies dare not claim they will decrease their prices because increasing their market share will harm their competitors. It is a world in which the government (almost) always wins, even if the courts must create markets that exist only on the pages of their opinions. It is a world in which the courts use progress to condemn mergers. Under such antitrust laws, the forces that create new businesses—replacing older ones with usually larger, more efficient competitors that lower costs and improve quality for their customers—are viewed with hostility, although they often benefit the less well-off the most."

In 2018, AT&T merged with TimeWarner after a legal challenge from the US Department of Justice. After testifying in support of the Department of Justice and opposing the merger, Carl Shapiro discusses "Vertical Mergers and Input Foreclosure Lessons from the AT&T/Time Warner Case" (*Review of Industrial Organization 2021*, pp. 303–341, https://link.springer.com/article/10.1007/s11151-021-09826-x). He writes: "I testified in court as the DOJ's economic expert in that case. I explain here how to quantify the increase in rivals' costs and the elimination of double marginalization that are caused by a vertical merger and how to evaluate their net effect on downstream customers. I also explain how this economic analysis fits into the three-step burden-shifting approach that the courts apply to mergers under Section 7 of the Clayton Act. Based on my experience in the *AT&T/Time Warner* case, I identify a number of shortcomings of the 2020 Vertical Merger Guidelines." After testifying in support of AT&T and in favor of the merger, Dennis W. Carlton, Georgi V. Giozov, Mark A. Israel, Allan L. Shampine provide "A Retrospective Analysis of the AT&T/Time Warner Merger" (*Journal of Law and Eco*nomics, November 2022, pp. S461–S497, https://www.journals.uchicago.edu/doi/10.1086/721268). "We describe and evaluate in detail the economic model used by the government's expert and then focus our empirical work on the accuracy of the predictions made by that model. We also discuss evidence related to the Comcast/NBC Universal merger, which involved the same theory of harm and was allowed to proceed with a remedy similar to the contractual commitment that AT&T/Time Warner unilaterally adopted. We conclude that the evidence from the time of trial showed the theory of harm to be weak and the specific empirical predictions made by the government's expert to be wrong. Postmerger evidence confirms that conclusion, as does new evidence from the earlier Comcast/NBC Universal merger."

Brian Albrecht, Dirk Auer, Eric Fruits, and Geoffrey A. Manne evaluate "Doomsday Mergers: A Retrospective Study of False Alarms" (International Center for Law and Economics, March 22, 2023, https://laweconcenter.org/resources/doomsday-mergers-a-retrospective-study-of-false-alarms/). "[T]his paper analyzes whether previous doomsday merger scenarios have materialized, or whether the critics' claims missed the mark. Our retrospective analysis shows that many of the alarmist predictions of the past were completely untethered from prevailing market realities, as well as far removed from the outcomes that emerged after the mergers." The six past mergers they discuss are Amazon-Whole Foods, ABI-SABMiller, Bayer-Monsanto, Google-Fitbit, Facebook-Instagram, and Ticketmaster-Live Nation.

Interviews

Yasheng Huang discusses "the development of the Chinese state" with Tyler Cowen ("Conversations with Tyler," March 8, 2023, https://conversationswithtyler.com/episodes/yasheng-huang/). On "a big misconception about China's economy," Huang says: "[O]ne of them is that they look at the Chinese R&D spending, and they look at, for example, some of the impressive technological progress the country has made, and then they drew the conclusion that the Chinese economy is driven by productivity and innovations. In fact, studies show that the total productivity contributions to the GDP have been declining in the last decade and even more. As China has begun to invest more in R&D, the economic contributions coming from technology, coming from productivity have been actually declining. In the economic sense, it's not a productivity-driven economy. It is an overwhelmingly investment-driven economy. I think that's one of the biggest misunderstandings of Chinese economy. It entails implications about the future prospects of the country, whether or not you can sustain this level of economic growth purely on the basis of massive investments." On China's leadership: "This is a remarkable statistic: Since 1976, there have been six leaders of the CCP [Chinese Communist Party]. Of these six leaders, five of them were managed either by Mao or by Deng Xiaoping. Essentially, the vast majority of the successions were handled by these two giants who had oversized charisma, oversized prestige, and unshakeable political capital. Now we have one leader who doesn't really have that. He relies mostly on formal power, and that's why he has accumulated so many titles, whereas he's making similar succession errors as the previous two leaders. . . . Xi Jinping does not match, even in a remote sense, the charisma and the prestige of Mao Zedong and Deng Xiaoping. There's no match there."

David A. Price interviews Annamaria Lusardi "on financial literacy, seniors versus scammers, and learning from the mistakes of NFL players" (*Econ Focus*: Federal Reserve Bank of Richmond, First Quarter 2023, pp. 24–28, https://www.richmondfed.org/publications/research/econ_focus/2023/q1_interview). Lusardi and Olivia Mitchell have designed a 28-question test to measure financial literacy, which has become a widely used research tool. Lusardi notes: "Together with a team at the World Bank, I eventually designed questions . . . that were applied to a sample of more than 140 countries. I would say there are several interesting findings. One is that even though the U.S. is the country with the most advanced financial markets, it actually doesn't score very high in terms of financial literacy. And this has been true in other surveys, as well. The second thing is that overall financial literacy is not high in other countries, either. Overall, the level of financial literacy globally is really low; only one-third of people around the world are financially literate. . . . [W]hat we did recently—and it took us a good many years to do this project— is a meta-analysis of financial education programs. . . . What we found, looking at the evidence in as many as 33 countries, is that financial education works and works well—meaning it does translate into higher knowledge and also better behavior in savings and managing credit and in other areas, including insurance and money transfers. And we also found that it is cost effective. This is due to the fact that many educational programs do not cost very much."

Daniel Kahneman tells Joseph Walker an origin story of "reference class forecasting" in in an interview on the Jolly Swagman podcast ("#143: Dyads, And Other Mysteries—Daniel Kahneman," April 14, 2023, https://josephnoelwalker.com/143-daniel-kahneman/). "Well, first let's define our terms, what the reference class is. I don't know a better way of doing this than telling the origin story of that idea in my experience, which is that, 50 years ago approximately, I was engaged in writing a textbook with a bunch of people at Hebrew University, a textbook for high school teaching of judgement and decision making. We were doing quite well, we thought we were making good progress. It occurred to me one day to ask the group how long it would take us to finish our job. There's a correct way of asking those questions. You have to be very specific and define exactly what you mean. In this case I said, 'Hand in a completed textbook to the Ministry of Education—when will that happen?' And we all did this. Another thing I did correctly, I asked everybody to do that independently, write their answer on a slip of paper, and we all did. And we were all between a year and a half and two and a half years. But one of us was an expert on curriculum. And I asked him, 'You know about other groups that are doing what we are doing. How did they fare? Can you imagine them at the state that we are at? How long did it take them to submit their book?' And he thought for a while, and in my story he blushed, but he stammered and he said, 'You know, in the first place they didn't all have a book at the end. About 40%, I would say, never finished. And those that finished . . .' He said, 'I can't think of any that finished in less than eight years—seven, eight years. Not many persisted more than ten.' Now, it's very clear when you have that story, that you have the same individual with two completely different views of the problem. And one is thinking about the problem as you normally do—thinking only of your problem. And the other is thinking of the problem as an instance of a class of similar problems. In the context of planning, this is called reference class planning."

Discussion Starters

Victor Matheson provides an overview of developments in "Sports Gambling" (*Milken Institute Review*, Second Quarter 2023, pp. 12–21, https://www.milkenreview.org/articles/sports-gambling). "The past 60 years have witnessed a massive transformation of the gambling landscape in the United States. In the early 1960s, the only legal casinos in the country operated in Nevada, no states ran lotteries, and essentially all sports bets were either made informally among friends or through illegal bookies. These days, 45 state governments sell $100 billion in lottery tickets each year, with multistate lotto games like Powerball and Mega Millions occasionally offering jackpots exceeding $1 billion. Nearly 1,000 casinos and card rooms operate across 41 states, generating over $50 billion in net gaming revenue. And nowhere has the gambling industry changed more rapidly than in sports betting, where nationwide expansion has led to an increase in legal wagering from just under $5 billion in 2017 to nearly $100 billion in 2022.

Michelle Clark Neely, "Is the Era of Overdraft Fees Over?" (*Regional Economist*, Federal Reserve Bank of St. Louis, March 8, 2023, https://www.stlouisfed.org/publications/regional-economist/2023/mar/is-era-overdraft-fees-over). "A relatively small proportion of bank customers account for the lion's share of overdraft fees. According to the Consumer Financial Protection Bureau (CFPB), people who frequently overdraft their accounts represented just 9% of bank customers but generated almost 80% of overdraft and nonsufficient funds (NSF) fees in 2017. Consulting firm Oliver Wyman estimates that customers who heavily use overdraft services generate, on average, more than $700 in profit for the bank per year on a basic bank account; customers who don't use overdraft services produce an average of $57 in profit for the bank per year. . . . [C]onsumer backlash, scrutiny from Congress and regulators, and increased competition from nonbank providers have prompted banks—especially large ones—to change or eliminate their overdraft programs. A growing number of banks are offering lower-cost alternatives to overdraft services that can cover the account deficit and help customers fulfill other goals, such as building credit."

Finance & Development has published some recent essays on industrial policy, including Ruchir Agarwal on "Industrial Policy and the Growth Strategy Trilemma" (March 21, 2023, https://www.imf.org/en/Publications/fandd/issues/Series/Analytical-Series/industrial-policy-and-the-growth-strategy-trilemma-ruchir-agarwal) and Douglas Irwin on "The Return of Industrial Policy" (June 2023, https://www.imf.org/en/Publications/fandd/issues/2023/06/the-return-of-industrial-policy-douglas-irwin). From Agarwal's essay: "Former US Treasury Secretary Lawrence Summers recently said he liked his industrial policy advisers the same way he liked generals. 'The best generals are the ones who hate war the most but are willing to fight when needed. What I worry is that people who do industrial policy love doing industrial policy.' . . . Just like salt in cooking, a pinch of industrial policy can be helpful, but too much can overpower, and prolonged excess can harm."

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

Summer 2023, Volume 37, Number 3

Symposia

AMERICAN ECONOMIC ASSOCIATION