The Journal of Economic Perspectives

A journal of the American Economic Association

Winter 2026

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.

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.
Printed at Sheridan Press, Hanover, Pennsylvania, USA.

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Winter 2026, Vol. 40, No. 1. The JEP is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. For details and further information on the AEA go to https://www.aeaweb.org/. Periodicals postage paid at Nashville, TN, and at additional mailing offices.

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

The Journal of
Economic Perspectives

Contents

Volume 40 • Number 1 • Winter 2026

Symposia

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 40, Number 1—Winter 2026—Pages 3–26

The Likelihood of Persistently Low Global Fertility

Michael Geruso and Dean Spears

Fertility is low or falling across the world: among high-, middle-, and low-income countries; among secular and religious populations; and in economies where the state is large and where it is small. Birth rates have been falling not only for decades, but for centuries. They have been falling for as long as there are good historical records to document them. Occasional bumps—even big ones like the mid-twentieth-century Baby Boom—have been temporary variations around this long-term global trend.

Even as population sizes have been soaring upwards, from a global one billion around 1800 to over eight billion today, birth rates have been declining. The global population has grown over recent centuries mainly because of falling mortality rates—particularly child mortality rates (Notestein 1945; Fogel 2004; Deaton 2013). Mortality rates fell earlier and faster than fertility rates, so the global population climbed rapidly, despite that family sizes were becoming smaller. The latest continuation of this long trend has attracted new attention from researchers, policymakers, and the general public.

A core summary measure of birth rates in a population is the total fertility rate (TFR), which captures the number of children that a representative woman would have over her lifetime if she followed the age-specific patterns of childbirth

■ *Michael Geruso and Dean Spears are both Associate Professors of Economics at the University of Texas at Austin. Geruso is also a Faculty Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Spears is also a Research Fellow, Institute of Labor Economics (IZA), Bonn, Germany, and Founding Executive Director of r. i. c. e., Uttar Pradesh, India. Their email addresses are mike.geruso@utexas.edu and dspears@utexas.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251463.

in a particular place and time (usually a calendar year). The TFR has fallen from a global average that was a little under five in 1950 to a global average that is a little over two in 2025. If the fertility rate for the world as a whole falls below two and stays below two, then there would be fewer than two children in each next generation to replace two adults in each previous generation.

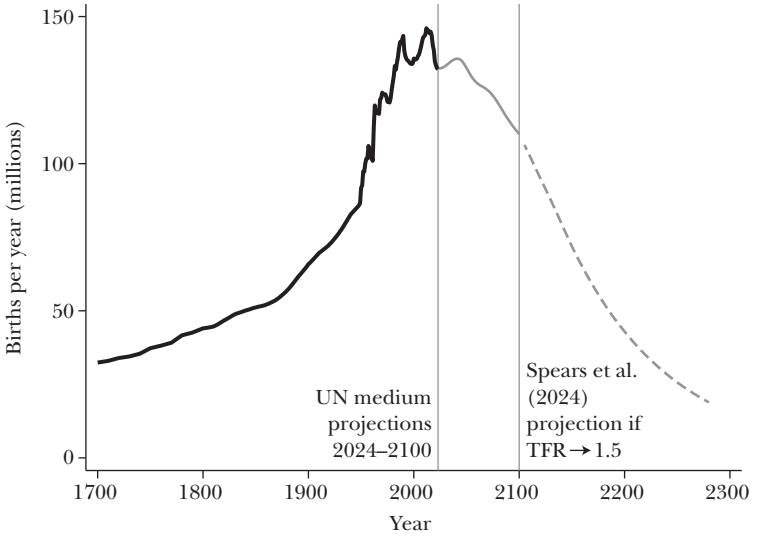
To understand the pace of potential change, consider a total fertility rate that held at 1.0, which is not far from the 2025 statistics for Chile, China, Puerto Rico, and Singapore. If in a first generation there were a hypothetical 100 people, then a fertility rate of 1.0 would imply that, in the second generation, these 100 would have 50 children. That is, an average of one child per woman—or alternatively an average of one child per two adults. In the third generation, those 50 would produce 25 children. In this way, progressing from a grandparent's generation to a grandchild's, the size of a generation would shrink by 75 percent. Even at birth rates greater than 1.0, the decline would be fast. The 115 richest countries in the world together have an average total fertility rate of 1.5 (United Nations 2024). A birth rate of 1.5 would lead to a decline of 44 percent in generation size over two generations.[1] With a birth rate of 0.7 children per two adults—approximately South Korea's 2025 total fertility rate—a similar calculation would yield 12 grandchildren per 100 people in a grandparents' generation.

In some large countries and regions—notably, China, Russia, and Europe—populations have already begun shrinking. Over the long run and globally, a worldwide average birth rate below two would imply that global birth cohorts—meaning, the groups of people born each year—would shrink decade by decade. In another article in this symposium, Lant Pritchett discusses the possibilities and challenges for labor mobility across borders to ease the strains of population aging and a declining share of the population in the labor force. Here we consider the world as a whole. From that perspective, immigration merely changes the geographic distribution of the global population, not its size.

Global change could come rapidly if birth rates remain low. In Figure 1, we plot estimates and projections of the annual count of worldwide births, tracing back several hundred years into the past and projecting over 200 years into the future. Projections from the present to 2100 are the "Medium" projection of the 2024 UN World Population Prospects (United Nations 2024). Beyond 2100 (where the UN projection ends), Spears et al. (2024) generate a cohort-component projection for a scenario in which the global average fertility rate converges to 1.5. There is no reason to expect convergence to exactly 1.5 (or to any other number, in particular); this hypothetical merely illustrates the pace of population change in a scenario in

[1] The estimate of 44 percent equals the decline after two generations of shrinking at 25 percent per generation: $1 - (1.5/2.0)^2$. For simplicity, the calculations in this paragraph assume a replacement rate of 2.0, which is slightly lower than any true replacement rate, as we discuss in detail in the next section. Substituting a higher replacement rate assumption, such as 2.1 children per woman, would yield faster generational declines.

Figure 1
The Peak in Annual Births Is in the Past



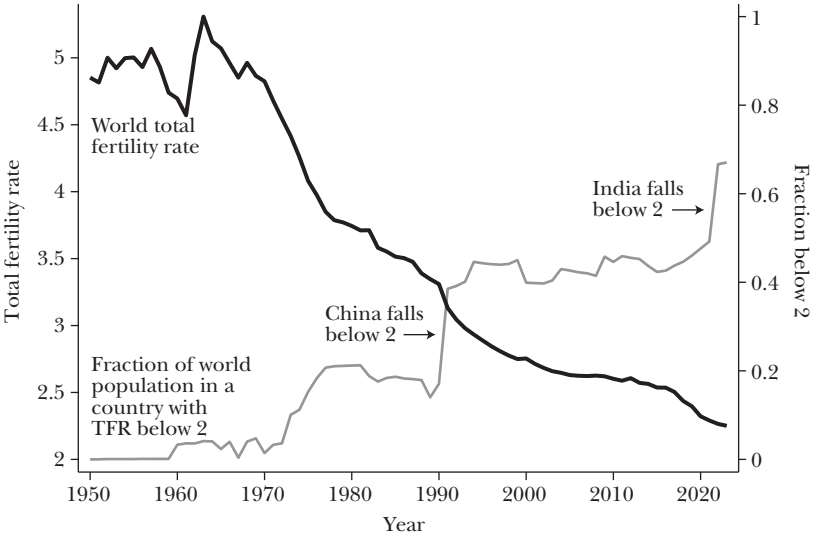
Source: Historical data through 1949 are replicated from Spears et al. (2024) (see Weston 2024), which itself draws on sources compiled by Our World in Data (Mathieu and Rodés-Guirao 2022). Data from 1950 to 2100 are estimates and projections from the 2024 UN World Population Prospects (United Nations 2024). Beyond 2100 are cohort-component projections from Spears et al. (2024) that extend the UN Medium projection in a scenario in which the global average total fertility rate converges to 1.5.
Note: The figure displays the worldwide count of annual births.

which fertility remains low where it is low today and continues falling where it is falling today.[2]

At the start of Figure 1, in 1700, there were perhaps about 30 million births annually. For comparison, estimates from Kaneda and Haub (2022) indicate that there were perhaps 18 million births per year globally 2000 years ago. In 2012, at the peak, there were 146 million births worldwide. Annual births have been falling since. In the hypothetical scenario in which the world converged to an average total fertility rate of 1.5 and held at that level, then by 2300, annual global births would have fallen back down below 16 million. Thus, birth rates that are already common today in most of the world would lead to a decline in the number of global births that could be more rapid than the increase of recent centuries. Moreover, notice that the decline does not end where the graph does: As long as birth rates hold at any level below two children per two adults—even at some steady, unchanging level—the slide down would continue indefinitely.

[2]Spears and Geruso (2025) plot a related but distinct history and projection that focuses on population size rather than birth cohort size.

Figure 2
Global Fertility Has Been Declining to Below-Replacement Levels



Source: Data are from the UN World Population Prospects 2024 (United Nations 2024).
Note: Figure displays, for 1950 to 2023, the worldwide period total fertility rate (left axis) and the fraction of the world population residing in a country with a fertility rate below 2 (right axis). We follow the UN World Population Prospects convention of including UN member states as well as territories, regions, and other designations, such as Taiwan.

These projections raise the questions that are the focus of this article: Is this sort of future, in fact, likely? Could low birth rates persist? The facts and evidence that we present here indicate that sustained global depopulation is a likely possibility.

Fertility Trends in Periods and Cohorts

The Total Fertility Rate: Births in a Period

To shed light on the future of birth rates, we begin by reviewing the evidence on global birth rates over the past 75 years. The United Nations Population Division (United Nations 2024) has compiled country-level estimates of total fertility rates back to 1950. Drawing on these data, Figure 2 plots the total fertility rate as a global average since 1950. The total fertility rate is a measure that aggregates information on age-specific fertility rates in a given year—for example, the rate of births to women 20–24 years old in 2023, to women 25–29 years old in 2023, to women 30–34 years old in 2023, and so on—into a summary measure that describes fertility at that point in time. This construction yields the answer to the question of how many children a representative woman would have over a lifetime if she experienced exactly the average age-specific birth rates of some snapshot in time.

The total fertility rate, as a summary statistic describing a time period, is analogous to life expectancy, which aggregates the information contained in a set of time-period-specific and age-specific mortality rates. Just as life expectancy is not a forecast of the expected longevity of someone born today, the total fertility rate is not a forecast of the expected childbearing of someone born today (or entering adulthood today, or having a first child today). Instead, it offers a readily-calculable summary of age-specific fertility rates during some period—typically a calendar year.

Figure 2 shows that the global average of the total fertility rate has been falling for most of the past 75 years: from 4.85 in 1950 to 2.25 in 2023 (the most recent estimate). Although there is important geographic heterogeneity, all regions of the world have experienced fertility declines over this long window: Europe fell from 2.7 to 1.4, North America from 3.1 to 1.6, and South America from 5.6 to 1.7. The two most populous countries also declined dramatically. China's TFR declined from 5.8 to 1.0. India—once the center of fear-mongering about overpopulation (for example, Ehrlich 1968)—fell from 5.7 to a little below 2.0. Sub-Saharan Africa, where fertility today remains highest among any region in the world, has fallen to 4.3 in the most recent estimate, after remaining high and largely stable (in a range from 6.3 to 6.8) from 1950 to 1990.

The apparent trend break in 1970, where the total fertility rate begins falling in Figure 2, is in fact a resumption of a much longer-term decline. Estimates of a global average fertility rate extend back decades, not centuries. But for individual countries with longer histories of reliable data, fertility by 1950 was far below its 1850 levels (for examples from Europe, see Doepke et al. 2023). Coale (1986, Figure 1.4) reports that the total fertility rate in France fell from about four in 1800 to below three by 1910 and in Sweden from above four in 1800 to near two in 1930. Temporary bumps like the mid-twentieth-century Baby Boom—the peak and end of which is captured in Figure 2—have been fluctuations around this long-term trend. Indeed, even during the year when the world's population growth rate was at its highest, 1968, global fertility was falling.

Fertility rates have fallen below two in many countries. A second trace in Figure 2 shows that in 1950, almost nobody lived in a country with a fertility rate below two. In 1975, about 15 percent of people did. By 2000, 40 percent of people did. And in the most recent estimates (for 2023), 67 percent of people in the world lived in a country with a fertility rate below two. Each of the three most populous countries—India, China, and the United States—is now below this threshold.

The threshold of two is important because it means that fertility is below the "replacement level." Below-replacement fertility at a national level does not necessarily imply a shrinking national population: National populations can grow, despite below-replacement fertility, due to migration. Even absent positive net migration, population sizes can continue increasing for a short period after birth rates fall below the two-for-two replacement rate because of improvements in mortality or because of the echoes of large cohorts from past decades. But over the long-term, populations with below-replacement fertility would shrink absent net in-migration.

Replacement-level fertility is not exactly 2.0. Instead, it is a variable that depends on context-specific facts about mortality and must account for the slightly unequal proportion of males and females born (Espenshade, Guzman, and Westoff 2003; Gietel-Basten and Scherbov 2019). Despite this complexity, it is common to read that replacement-level fertility is 2.1.[3] Here, we sometimes simplify the discussion by discussing "two" as a focal threshold. Although the exact replacement rate varies by context, if fertility rates are below 2.0, then they are below the replacement rate specific to any time and place.

Completed Cohort Fertility: Births to a Cohort

The total fertility rate is a useful summary statistic, but it does not perfectly capture whether a population will grow or shrink over the long term. That is because TFR is a period rate—it depends only on what happens in a given year. Long-term population change depends on how many children parents have over their lifetimes. Period fertility can change without lifetime fertility changing if the ages of childbearing shift.

The ages of childbearing have indeed been changing. Since the mid-twentieth century, the average age at motherhood has been increasing in many places (Lesthaeghe 1995). In the United States, the mean age of childbearing rose from 24.6 in 1970 to 27.5 in 2023 (Mathews and Hamilton 2002; Brown et al. 2025). Birth rates at earlier ages have been declining, and birth rates at later ages have been increasing, especially births to women above 35 in low-fertility countries (Sobotka and Beaujouan 2017). Such changes can complicate the interpretation of period fertility rates, like those shown in Figure 2. As a simplified illustration, consider a shift from a scenario in which a mother in an earlier cohort had a first child at 25 and a second child at 27 to a scenario in which a mother from a later cohort had a first child at 34 and a second child at 38. An effect of this relative delay and longer birth spacing among later cohorts would be that period total fertility rates would first decline and then increase, yet such a shift would have no impact on the number of children born to each family.

Period total fertility rates are sensitive to such "tempo" shifts—meaning shifts in the age-patterns of childbearing—because these measures combine age-specific fertility rates of different birth cohorts at a point in time. For example, the TFR in France in 1990 would combine information on age-35 fertility among women born in 1955 with information on age-36 fertility among women born in 1954, and so on. Completed cohort fertility (CCF), in contrast, captures lifetime births to the cohort of women born in some place and time—for example, the average number of children born to women who themselves were born in France in 1950. Therefore, to get

[3]The replacement rate (or replacement TFR) is not 2.1 but covaries with female mortality, especially child female mortality. If female mortality before age 50 were zero, replacement TFR would be about 2.05 because, absent sex-selective abortion, about 105 males will be born for every 100 females. Therefore, 100 females reproducing 100 females would imply about 205 births, or 2.05 births per woman. The replacement rate rises with female mortality, as not all females will survive through all childbearing years. Male mortality does not enter the calculation.

a fuller picture of the long term trends in declining fertility, it is useful to examine CCF as well.[4]

The data requirements necessary to calculate completed cohort fertility are more exacting, typically requiring either a representative sample of retrospective birth histories collected after childbearing is completed or detailed information on age-specific fertility rates covering all of the years in which the cohort was aged 15–49. For some countries, this information is available: The Human Fertility Database aggregates historical information on cohort fertility rates for a set of relatively rich countries for which high-quality vital statistics data have been consistently collected (Human Fertility Database 2023).

In Figure 3, we compare period total fertility rates from the UN World Population Prospects with completed cohort fertility data from the Human Fertility Database for a select group of countries. Each panel is a country. To compare period and cohort rates on a common timescale, cohort data are plotted in the figure according to the year in which that cohort turned 30. We plot the maximal extent of available country-specific data on cohort rates from the Human Fertility Database, alongside the total fertility rate data from 1960 to 2023. Figure 3 shows that CCF has followed a downward trend that is similar, though not identical, to period TFR. For the countries and time periods available, CCF has tended to lag TFR and to be less steeply sloped. The time span over which there is overlapping support for the two measures was a period during which TFR was falling and average ages at birth were rising. In all cases, CCF has fallen, though the United States presents a contrast to the global pattern over these years, a point we return to below. Both TFR and CCF in the United States were fairly stable, and even rose slightly, for three decades spanning 1980 to 2010.

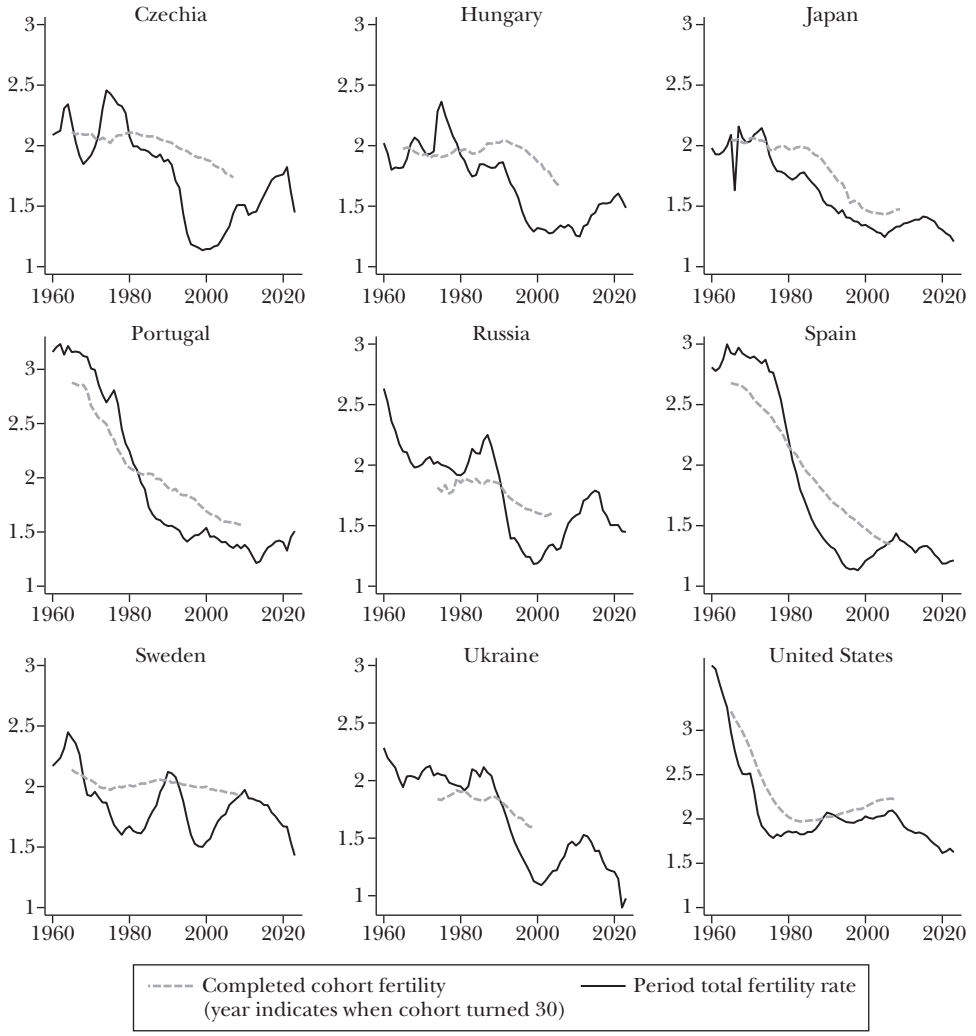
Completed cohort fertility tends to evolve more smoothly than period fertility rates, so the two measures can diverge significantly over the short run of a decade or two—as the panels of Czechia, Sweden, and Russia show. This contrast between CCF and TFR reveals something critical in assessing policies that may impact fertility: If a policy change causes some mothers to choose childbearing earlier than they otherwise would have, but does not affect the total children ever born to those women, then it will influence TFR and yet have no impact on the outcomes that matter for long-run population trends—and no influence on CCF. We return to this point later in our discussion of how policies have been shown to affect fertility.

Child-less or Child-fewer?

Recent popular and research attention has focused on the role of childlessness in accounting for patterns of low fertility in the United States and around

[4]Whereas a period TFR is calculated as $TFR = \sum_{age=15}^{49} ASFR_{age,\,calendar\ year}$, completed cohort fertility could be calculated as $CCF = \sum_{age=15}^{49} ASFR_{age,\,cohort}$, where $ASFR$ is an age-specific fertility rate calculated with data from a calendar year or birth cohort, respectively.

Figure 3
Period Total Fertility Often Varies More Than Completed Cohort Fertility



Source: TFR data are from the UN World Population Prospects 2024 (United Nations 2024). Completed cohort fertility data are from the Human Fertility Database (Human Fertility Database 2023).
Note: Figure displays period total fertility rates (TFRs) and completed cohort fertility (CCF) for a select group of countries where sufficient CCF data were available. Cohorts are plotted at their date of birth plus 30 years. For example, the CCF for the 1970 birth cohort is plotted at 2000. We plot the full range of available cohort years from the HFD, which differ by country.

the world (for example, Stone 2020; Kearney, Levine, and Pardue 2022; Doepke et al. 2023; Wolfe 2024; Lewis-Kraus 2025; Ermisch 2025). Both the total fertility rate and completed cohort fertility are averages, and so they are not informative on the question of whether falling birth rates have been driven primarily by changes in

the probability of ever giving birth. They do not distinguish, for example, between a scenario in which two women in the age range 30 to 34 each had one child and a scenario in which one of these women had two children over these ages and the other had none.

Characterizing childlessness matters because long-term depopulation may be especially plausible if both high rates of childlessness and low birth rates among parents are each contributing factors. In this section, we decompose changes in cohort fertility into two components: an increase in childlessness (lifetime nulliparity) versus fewer children being born among women who do have children (completed cohort fertility among the parous subcohort).[5]

We focus, necessarily, on the countries and cohorts for which high-quality parity data are available. We use two datasets. The first is all of the country-cohorts with available childlessness data in the Human Fertility Database. This includes 683 cohorts of women from 25 countries in eastern and western Europe, North America, and Asia.[6]

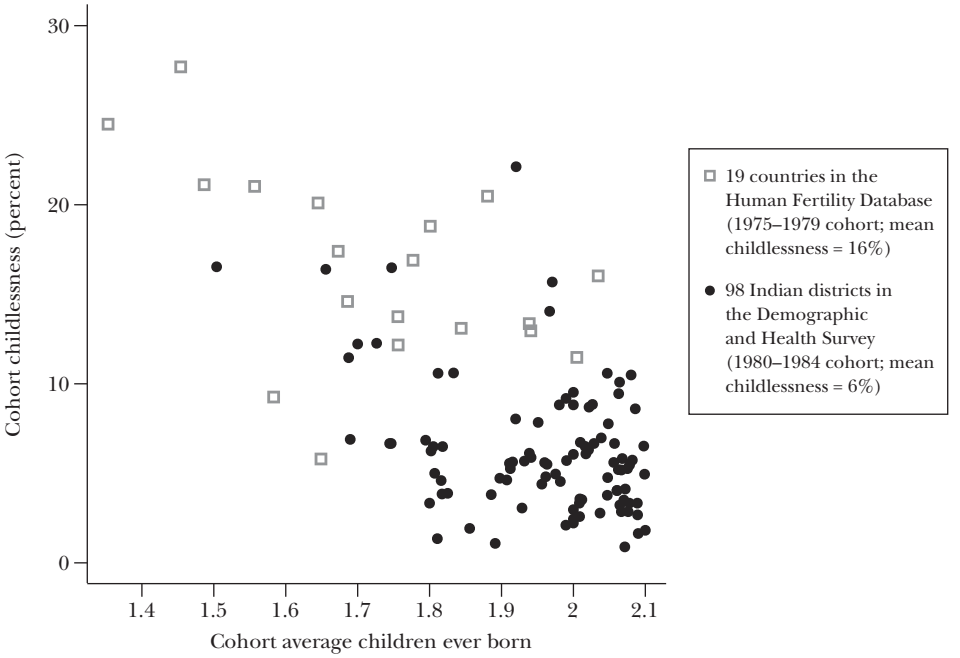
To examine this question in the context of low- and middle-income populations, we also draw on a second sample: birth histories in India's most recent Demographic and Health Survey (2019–2021), known as the NFHS-5 in India, which collects data on children ever born from interviews of women aged 15 to 49 (IIPS and ICF 2021). The survey is representative of what has become the largest national population on Earth, about 1.4 billion people. India's emerging status as a below-replacement population has attracted recent notice: for useful starting points, see Gietel-Basten and Scherbov (2019), Gietel-Basten, Spears, and Visaria (2022), Visaria (2022), and Park et al. (2023), as well as the extended discussion in Spears and Geruso (2025). However, it remains uncommon to include India in comparative studies of low fertility. We examine geographic heterogeneity in fertility across Indian districts, which are administrative subdivisions of states. The average Indian district population size in 2021 was about two million people, which would be comparable to the country population of Slovenia or the size of a small US state, such as Idaho or Nebraska.

Our analysis of data from India focuses on a single pooled cohort (born 1980–1984) and draws insights by comparing higher- and lower-fertility districts within India. We compute cohort fertility among the parous (that is, those who have given birth) as a simple sample mean, conditional on ever having given birth at time of survey. These women would have been observed at ages 35 to 41. This restriction

[5] This section of the paper subsumes and abridges the content of our working paper "Childless or Childfewer? Childlessness and Parity Progression where Fertility is Below Replacement" (Geruso and Spears 2025). Contemporaneously, Ermisch (2025) conducted a complementary analysis of childlessness in England and similarly found that both factors contribute to cohort fertility decline.

[6] Cohorts in the Human Fertility Database span the birth years 1935 to 1979. We drop subunits that are tallied separately within the database—East and West Germany, and England and Wales, Scotland, and Northern Ireland—while keeping Germany and the United Kingdom. This leaves 25 countries: Austria, Belarus, Bulgaria, Canada, Czechia, Denmark, Estonia, Finland, Hungary, Iceland, Ireland, Japan, Lithuania, Netherlands, Norway, Poland, Portugal, Russia, Slovakia, Slovenia, Spain, Sweden, Taiwan, Ukraine, and the United States.

Figure 4
Cohort Childlessness Varies Widely Conditional on Cohort Average Fertility



Source: Human Fertility Database (2023) and IIPS and ICF (2021).
Note: The sample includes Human Fertility Database (HFD) countries for which there is cohort childlessness data and Indian districts. Cohort childlessness for Indian DHS survey respondents is defined as reporting no children at the time of survey. Both subsamples are restricted to cohorts with cohort fertility below 2.1.

balances a desire to focus attention on the most recent cohorts against the limitation that some survey respondents will not have completed their lifetime childbearing at the time of survey.[7] The Human Fertility Database, in contrast, measures completed cohort fertility and includes both cross-sectional variation and variation over time, with dozens of birth cohorts per country.

We begin by showing geographic heterogeneity in childlessness and cohort fertility. In Figure 4, each data point is an Indian district or the most recent country cohort from an advanced economy in the Human Fertility Database.[8] Restricting

[7]That limitation is less quantitatively important in the modern Indian context than, for example, in the United States, because most marriages and births in India happen relatively young. It is especially immaterial given our focus is on the role of childlessness, because these data reveal that childlessness in India by age 35 is relatively rare: less than 4 percent overall for the 1980–1984 cohort and only 6 percent when conditioning on low-fertility Indian districts where cohort fertility is below 2.0.
[8]For convenience, we sometimes refer to the countries in the HFD or subsets of them as "advanced economies," but the International Monetary Fund (2025) classifies Belarus, Bulgaria, Hungary, Poland, Ukraine, and Russia as "emerging and developing."

attention to geographies with cohort fertility below 2.1, the figure shows that lower fertility is associated with a greater cohort share that is childless. But the figure also shows that there is significant variation in childlessness for a given level of cohort fertility. In other words, two places with the same cohort fertility might show very different fractions childless. In some low-fertility cohorts, childlessness is common—in excess of 25 percent. In others, childlessness is below 5 percent. There is a clear, visible difference in the figure between low-fertility populations in India and low-fertility cohorts in advanced economies elsewhere. Even at the same overall cohort birth rate, childlessness is less common in India than in the advanced economies in the plot (although in some Indian districts, cohort childlessness exceeds 15 percent).[9]

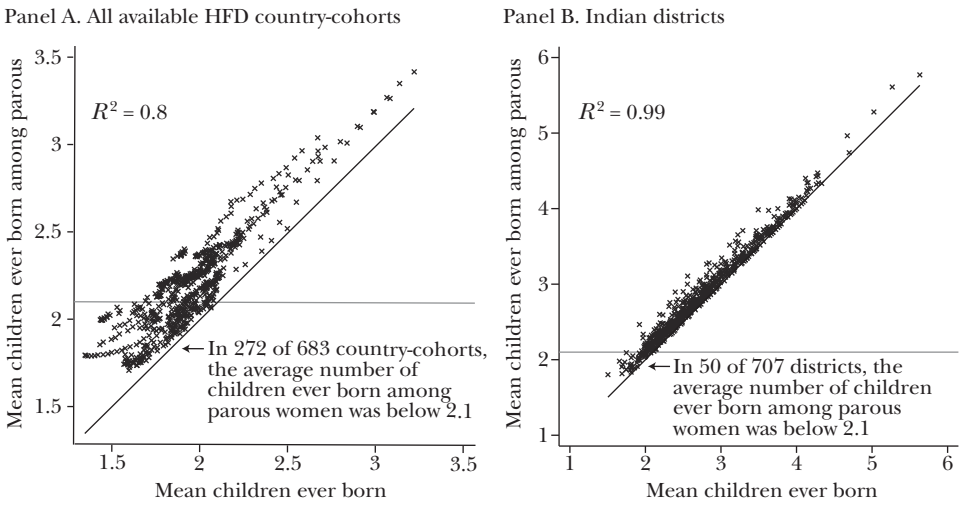
A complementary approach to understanding the role of childlessness in overall completed fertility is to focus on the average lifetime fertility among parous women (again, the nonchildless). Figure 5 compares this statistic, which excludes nulliparous women, against simple cohort fertility, which includes nulliparous women. In panel A, each data point is a cohort from an advanced economy.[10] In panel B, each data point is the cohort outcome from an Indian district. If there were no childlessness, all points would lie on a 45-degree line. In the other extreme, if all variation in completed cohort fertility arose from difference in childlessness, points would lie along the same horizontal line.

As Figure 5 shows, average cohort fertility among parous women is highly correlated with unrestricted completed cohort fertility for both country-cohorts and Indian districts. Most of the variation across the country cohorts in panel A—which in this figure includes cross-country comparisons and comparisons within countries over time—is explained by fertility among the parous ($R^2 = 0.80$), which means that it is not explained by childlessness. Childlessness explains even less in India: In panel B, the variation in childlessness explains almost none of the variation in cohort fertility across 707 Indian districts in the 1980–1984 cohort. There, fertility among the parous explains essentially all of the variation ($R^2 = 0.99$). In panel A, despite data points further from the 45-degree line, indicating a larger role of childlessness in determining overall birth rates, there are many country-cohorts for which the average number of children ever born among parous women was below 2.1: 272 of 683 country-cohorts fell below this threshold. In these places, cohort fertility would be below replacement even in a counterfactual thought experiment in which nulliparous women instead

[9]Formally, in a regression of childlessness on cohort fertility using the data points in the figure, we estimate a large and statistically significant effect on an indicator for being an Indian district cohort (rather than an HFD country cohort): At the same level of cohort fertility, Indian observations have 7 percentage points (standard error 1 percentage point) lower cohort childlessness. The estimating equation is $\textit{fraction childless} = \beta_0 + \beta_1\ \textit{cohort fertility}_i + \beta_2\ \mathbf{1}[\textit{India}_i] + \epsilon_i$, where cohorts are denoted by i, $\textit{India}_i$ is an indicator for an Indian cohort, and β_2 is the coefficient of interest. The regression is available in Geruso and Spears (2025b).

[10]Each country is represented multiple times in the figure. Average lifetime fertility among the parous in the HFD country-cohorts is calculated as completed cohort fertility divided by one minus the fraction childless. In the Indian DHS microdata, it is calculated as the simple mean of children ever born among women with one or more births.

Figure 5
Below-Replacement Average Cohort Fertility Is Found Even within Parous Subcohorts



Source: Data in panel A are from the Human Fertility Database (HFD); data in panel B are from India's 2019–2021 Demographic and Health Survey.
Note: For a list of countries included in panel A, see text discussion. In panel B each observation is an Indian district, which is an administrative subdivision of a state. Data points exactly on the diagonal 45° lines would reflect zero childlessness.

had birth rates that exactly matched their parous cohort peers. Even excluding childless cohort members, all of the available birth cohorts in Russia and Spain, and many in Japan, Portugal, and Taiwan, have below-replacement fertility.

To more precisely quantify the role of childlessness in accounting for below-replacement fertility, we conduct a decomposition in the spirit of Kitagawa's (1955) exact decomposition of a weighted average of subgroup means into the component due to weights and the component due to within-subgroup means. In our application, the two subgroups are the parous and nulliparous. The within-subgroup means are average cohort fertility among the parous and average cohort fertility among the nulliparous (which is zero by definition). The aim of the decomposition is to compare two cohorts, such as an earlier and later cohort of the same country, and evaluate the relative importance of increasing childlessness versus a change in the fertility rate among the parous in accounting for overall cohort fertility decline across the cohorts.

Table 1 reports the results, using data from the advanced economies in the Human Fertility Database. As the top panel shows, the average early cohort is born in 1956, while the average later cohort is born in 1976—that is, just old enough for members to have completed, or nearly completed, their childbearing years by 2025. The bottom panel shows the results of the decomposition. The average

Table 1

Decomposition of Fertility Decline: Childlessness versus Fewer Children among the Parous in Advanced Economies

	Most recent cohort	Cohort 20 years before
Panel A. Sample means		
Country-cohorts in sample	19	19
Average cohort year	1976.1	1956.1
Average completed fertility	1.77	2.01
Average completed fertility among parous	2.06	2.23
Average childlessness	0.14	0.10
Panel B. Kitagawa-like decomposition		
Total difference in completed cohort fertility:	0.25 births	
Difference in fertility among parous:	0.18 births	
Difference in childlessness:	4.4 percentage points	
Difference due to childlessness:	0.09 births, or 37% of gap	
Difference due to fertility among parous:	0.16 births, or 63% of gap	

Note: Decomposition compares cohorts separated by 20 years in 19 countries in the Human Fertility Database. Each observation is a country-cohort in the HFD, unweighted. The sample is smaller than in Figure 5 because a country must have two cohort-observations 20 years apart to be included. Some fractions may not add to one due to rounding. Full calculations are available in Geruso and Spears (2025b), which also lists the countries and cohorts included.

decline in fertility among these recent cohorts relative to the cohorts preceding them by 20 years was 0.25 births. Of this decline, 0.09 births, or 37 percent of the gap, is statistically accounted for by increased childlessness in the later cohort. The remaining 0.16 births, or 63 percent of the gap, is accounted for by declines in fertility among the parous.

A similar analysis can be used to decompose differences across districts in India, where the difference to be decomposed is across districts for women born in the same set of years, with two groups of districts defined by having the lowest and highest cohort fertility rates. Unsurprisingly, given panel B of Figure 5, almost all of this difference—94 percent—is accounted for by the difference in fertility among the parous. Differing patterns of childlessness account for only 6 percent of the gap between high-fertility and low-fertility districts.

These results complement a prior literature decomposing changes in cohort fertility with attention to parity within countries of Pacific Asia, Europe, and the United States (Zeman et al. 2018; Gietel-Basten 2019; Beaujouan, Zeman, and Nathan 2023). Other papers study childlessness or decompose fertility changes by parity with a smaller geographic focus on one or a few countries. These include Ermisch (2025), which finds that changes in childlessness were important to changes in English cohort fertility earlier in the twentieth century; Hellstrand et al. (2021) and Hellstrand, Nisén, and Myrskylä (2022), which study Nordic Europe; Hwang (2023), a study of South Korea; Sobotka (2021) on East Asia; and Aaronson, Lange, and Mazumder (2014) on Black women in the US South.

Our finding that changes in childlessness are an important—but not the primary—driver of falling fertility in many country contexts provides an informative contrast with recent US-centered narratives that attribute below-replacement birth rates primarily to increasing childlessness (as one example, Stone 2020). Childlessness is indeed more important to recent changes in cohort fertility in the United States than in the other advanced economies we examine here. The United States is somewhat of an outlier in cohort fertility trends, as shown above in Figure 3. Further, Kearney and Levine (2025) show that the most recent cohorts of US women are meaningfully more likely to be childless at 25 and 30 than earlier cohorts.

Overall, both increasing childlessness and smaller families among the parous are important—even though the contribution of each factor differs across England, India, the United States, and elsewhere. Fertility among parents in many countries is already low enough to lead to depopulation, even without childlessness. In short, sub-replacement fertility results from two strengthening trends: both greater childlessness and smaller family sizes among people who are parents.

Prospects for a Reversal

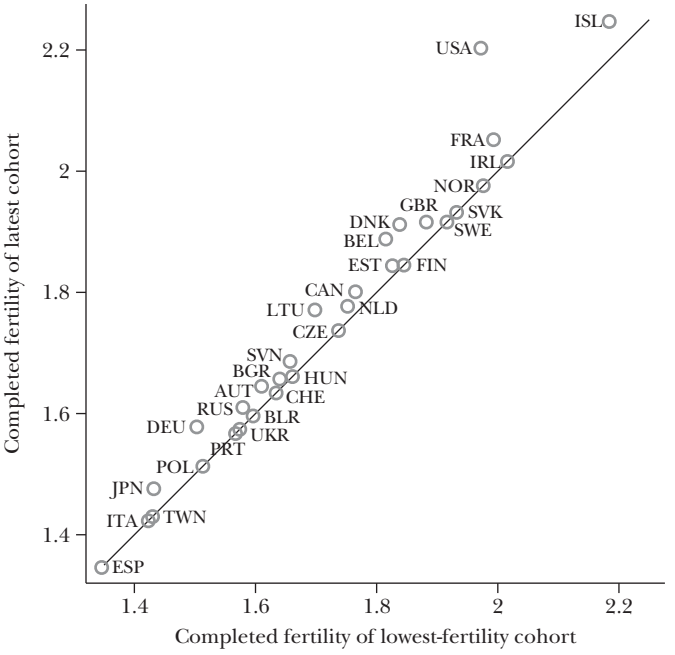
We opened this article by considering a scenario for the future of the global population if birth rates remained low where they are low and continued falling where they have been falling. What are the prospects for a reversal? Answering this question is complicated by the fact that below-replacement fertility—for any extended time, in any large national or subnational population—is a relatively new phenomenon, and therefore evidence on possible reversals is limited. Yet, we can examine what evidence there is.

Reversals to Date

A total fertility rate that is substantially below replacement is new to the United States, emerging over the past 15 years. But other countries crossed the threshold a half-century ago. Figure 2 showed that already by 1980, more than one in five people worldwide lived in a country with a TFR below two. Europe overall, as well as Japan, Australia, and Cuba, fell below two in the mid-1970s. These countries have not, so far, experienced a rebound to above-replacement levels. In terms of cohort fertility, where it is observable among the advanced economies represented in the Human Fertility Database, there have been 24 countries in which cohort fertility ever fell below 1.9. In none of these cases have subsequent cohorts from the same country ever had fertility as high as 2.1.[11]

[11] This 0-for-24 statistic is not a claim that no substantial increase in cohort fertility has ever happened in any time or place. Completed cohort fertility in the United States rose, for example, from 2.5 for the cohort of women born in 1900 to 2.9 for the cohort of women born in 1925 (US Department of Health, Education, and Welfare 1973). Ours is a claim about more recent decades, recognizing that the factors that shape the decision to have two children rather than three or four are likely different from the factors that shape the decision to have one or zero rather than two. The Human Fertility Database includes East

Figure 6
Completed Cohort Fertility Is at or near All-Time Lows



Source: Data are from the Human Fertility Database (HFD).
Note: Each datapoint is a country.

Beyond the stark binary of whether a rebound to replacement fertility has ever occurred, what about the question of whether cohort birth rates are rebounding to levels that are at least higher than their historical lows? In Figure 6, we compare the latest cohort to the lowest-fertility cohort from each country. The latest cohorts in the figure are born in 1979 or earlier, and so are at least 45 by 2025. Dots in the figure exactly along the 45-degree line would indicate that the most recent cohort with completed fertility data is also the lowest-fertility cohort on record. This is the case for 15 of the 31 countries in the figure. This distance from the 45-degree line to any dot above it indicates the extent to which rates have dipped and then recovered. In nearly all cases where the most recent cohort is not the lowest fertility cohort, fertility nonetheless remains close to the lowest historical value. In fact, if one were in search of a good predictor of the most recent cohort's fertility, simply predicting it with the fertility of the lowest-fertility cohort ever recorded in the country would be an excellent guide: The lowest cohort fertility within each country explains nearly

Germany, West Germany, Northern Ireland, Scotland, and together England and Wales as countries. If we count these separately (removing the United Kingdom and Germany as wholes) it is 0-for-26 instead of 0-for-24.

all of the cross-country variation in recent fertility rates ($R^2 = 0.96$). This pattern is not evidence that a rebound to replacement fertility could never occur, but it does suggest that a long-term, global future of low fertility is enough of a possibility to warrant attention.

The United States is a notable outlier in Figure 6—both in terms of having recent cohort fertility that is furthest from its lowest level and in having higher cohort fertility than other countries in the sample—even though cohort fertility in the United States is low from a world historical perspective. The 1953 US cohort experienced the lowest completed cohort fertility to date (Human Fertility Database 2023; Kirmeyer and Hamilton 2011). From that low of 2.0, fertility among US cohorts rose slightly to 2.2 for women born through the 1970s. The most recent US fertility data, however, suggest a reversal to lower fertility among more recent cohorts. Since around 2010 (when women born after 1980 were 30 or younger), US fertility has been falling rapidly (in this journal, Kearney, Levine, and Pardue 2022), possibly representing a delayed US convergence to the pattern of lower fertility found in most other advanced economies.

But what about increases in cohort fertility that started from levels higher than those we consider: Have there ever been notable increases from levels well above two to even higher levels? The twentieth-century Baby Boom is one example. Period fertility began rising in the 1930s and 1940s in many Western countries, including the United States, both as women became more likely to marry and as birth rates rose within marriages. This persisted in cohort birth rates so, as Van Bavel and Reher (2013) summarize, "[C]ohort fertility started to rise in most Western countries among the generations who began having children between the end of the Great Depression and the conclusion of World War II."

Although there have been many studies evaluating hypothesized causes of the Baby Boom—for example, see Bailey and Collins (2011) on improvements in household technology, Dettling and Kearney (2025) on the role of expanded mortgage access and housing supply, Albanesi and Olivetti (2014) on improvements in maternal mortality, and Doepke, Hazan, and Maoz (2015) on the role of World War II—there is no consensus on the relative contribution of these and other factors. The historical episode, despite its obvious importance, remains somewhat of a mystery from the perspective of social science. As a result, it is also an open question how much can be extrapolated from this episode (a temporary increase from above-replacement levels of fertility to even higher levels) that would speak to the possibility of future reversals from below-replacement starting points. And a narrower question also remains open: What sort of deliberate policy interventions (as opposed to broad but undirected social, technological, and economic changes) might be effective at generating a reversal?

Microeconometric Evidence and the Challenge of Identification

Another approach to assessing the likelihood of a fertility rebound is to consider the evidence on what policies, programs, and social and economic changes have been shown to increase fertility in low-fertility societies. By 2015, there were

55 country governments with explicitly pro-natal policy goals (Sobotka, Matysiak, and Brzozowska 2019; United Nations 2015). In other countries with low birth rates, changes in economic or social conditions have occurred that could matter for family formation, such as rising incomes or changes that affected the supply or prices of housing, education, or childcare. But the clear-cut bottom line is that whatever impacts pro-natal policies and broader changes might have caused, none has caused low birth rates to reverse enduringly back to replacement levels.

What about more modest effects? Might pro-child, pro-parent, or pro-family policies raise birth rates on the margin or slow their decline, even if these effects fall short of generating replacement-level fertility? Most pro-fertility policies in advanced economies primarily involve spending money: on childcare subsidies or government-run childcare centers, on paid parental leave, on targeted tax credits, or on cash for families. Several recent reviews survey the evidence of the effects of government spending in its various forms (Sobotka, Matysiak, and Brzozowska 2019; Doepke et al. 2023; Gauthier and Gietel-Basten 2025; Kearney and Levine 2025; and in this journal, Olivetti and Petrongolo 2017). We refer the reader to these for an in-depth discussion. Although cross-country comparisons often find positive correlations between fertility and public spending on children and families, studies evaluating particular national or sub-national policy changes have less consistently estimated positive effects across policy contexts. Where effects are positive, they tend to be small. The evidence for the long-term fertility impacts of parental leave is similarly mixed. Kearney and Levine (2025) conclude: "[O]ur read of the relevant evidence is that efforts that incrementally affect the ability to combine market work and parenthood—namely paid leave and expanded access to childcare—have at best only a small effect on fertility outcomes."

The microeconometric evidence on what policies may impact fertility is underdeveloped relative to the importance of these questions. One difficulty that researchers face is the distinction between period and cohort birth rates that is at the heart of this paper. Family formation is not a single decision at any one point in time. Rather, it is a series of choices that evolve over the life course and influence one another. Period fertility rates, which measure only the outcomes of choices at a particular point in time, can therefore importantly diverge from completed cohort fertility, which reflects influences over the entire life course. The difficulty has been recognized for decades (Hajnal 1947) and is well-studied in the population science literature (Bongaarts and Feeney 1998).

The upshot of this challenge is that a large effect on period fertility could have no effect on cohort fertility. If a new, generous childcare tax credit, for example, causes a pair of parents to have two children when the mother is age 29 and 31 instead of 33 and 36, then there could be a statistically detectable effect on period birth rates, but there would be no effect on cohort birth rates—and no effect on generation-to-generation population size. Several in-depth recent reviews have summarized the evidence in the literature with thoughtful attention to this difficulty. Sobotka, Matysiak, and Brzozowska (2019) explain: "Large-scale expansions of family policies often have considerable short-term effects on fertility, leading to

temporary baby booms and giving a time-limited boost to the period Total Fertility Rate." In terms of effects on completed fertility, Gauthier and Gietel-Basten (2025) summarize the evidence as showing "a relatively weak direct link between family policy interventions and quantum changes in fertility in low-fertility settings."

Estimating the effects of policies on completed cohort fertility—and not (only) on period, age-specific birth rates—is therefore important for understanding long-run effects. But in practice, this means studying cohort effects of policies and social changes that happened many decades in the past. For example, Dettling and Kearney (2025) study how variation in home loan availability and housing supply in the 1930s and 1940s affected fertility. This is far enough in the past to observe CCF. In contrast, it would be impossible, as of 2025, to observe CCF for cohorts born in the mid-1980s or later. Looking backward to social or economic changes that are now many decades in the past raises an external validity challenge: Because the social and economic environment today is so different from the mid- or late twentieth century (not least, because the level of fertility is lower), it is complicated to apply past results to today's question of what may cause birth rates like 1.6 to rise to a level like 2.0.

A more subtle challenge arises as well. The past few decades have seen the rise of microeconometric identification strategies that isolate exactly one dimension of variation in order to cleanly trace an outcome to the effect of a specific cause. Difference-in-differences, regression discontinuities, and instrumental variable estimates of policy effects are often most convincing when making comparisons that rely on one way or another on sharp, shorter-term responses to changes in the policy environment. In other words, they often rely on timing. For example, how did annual birth rate outcomes change following the introduction of a paid leave policy, comparing more and less policy-exposed groups? If the result is that researchers provide sharp answers to questions about short-term period birth rates, then larger questions about cohort effects might remain unanswered—even if each study of period, age-specific outcomes is impeccable.

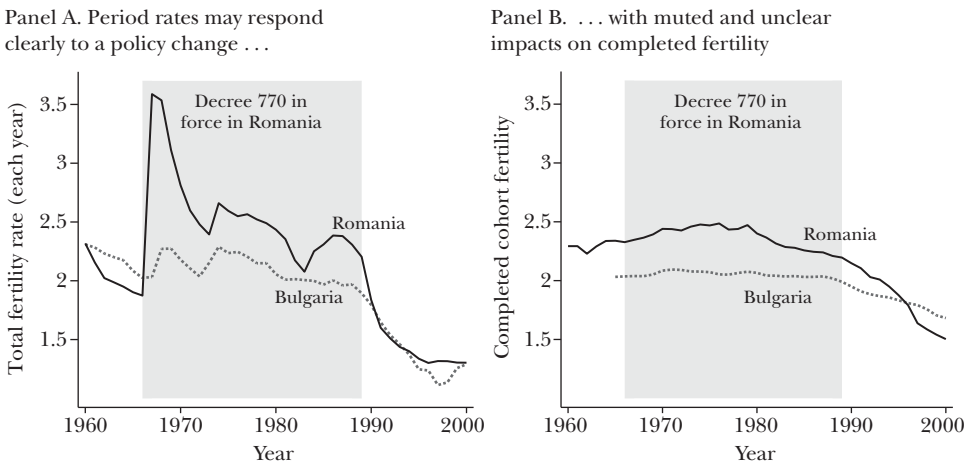
An Illustrative Example: Coercion in Romania

To illustrate this challenge of measurement and inference, Figure 7 presents Romanian birth rates before, during, and after the imposition of an infamously coercive policy aimed at raising births. In 1966, a dictatorial government imposed Decree 770, which banned abortion and made modern contraception effectively inaccessible. The figure extends an idea from Sobotka, Matysiak, and Brzozowska (2019), which compares cohort and period fertility rates in Romania over a similar evaluation window.[12] We add data from Bulgaria, Romania's neighbor that was also

[12] The figure is an attempt to replicate and build on Figure 18 from Sobotka, Matysiak, and Brzozowska (2019), using slightly different sources, and adding the comparison case of Bulgaria. For the Romanian total fertility rate, we use the same UN World Population Prospects data on period TFR used throughout this article (United Nations 2024), whereas the original figure draws on other sources. The Sobotka, Matysiak, and Brzozowska (2019) paper does not include the comparison to Bulgaria.

Figure 7

Period Rates Can Overstate Policy Impacts: Romanian Total Fertility Rate and Completed Cohort Fertility



Source: TFR estimates are from the 2024 UN World Population Prospects (United Nations 2024). Completed cohort fertility estimates for Bulgaria are from the HFD, and for Romania are from Gheţău (1997).

Note: Figure displays the period total fertility rates and completed cohort fertility for Romania. Cohorts are plotted at their date of birth plus 30 years. Decree 770 was in force from 1966 to 1989.

communist during the time of the policy and that might plausibly serve as a control, shedding light on what course Romanian fertility might have followed after 1967 if not for the policy. Panel A plots period birth rates in the two countries and shows that Romania and Bulgaria had substantially similar trends and levels in period total fertility rates before and after the Romanian policy window. Focusing on panel A of Figure 7, it is clear that birth rates in Romania changed dramatically following the start of the policy, as families were taken by surprise. TFR nearly doubled in the year that followed. The sharp timing of this apparent impact following the policy change, together with the availability of data from neighboring Bulgaria to serve as a control, suggests the possibility of a difference-in-differences analysis comparing birth rates pre– and post–Decree 770 in Romania and Bulgaria.

But while such an analysis could answer the narrow question of the causal effect of Decree 770 on the total fertility rate in 1967, it may nonetheless reveal little in terms of the impact of the policy on the number of children Romanian women had over their lifetimes. After the initial rise in TFR, birth rates soon began falling quickly in Romania, as behavior adapted to the new policy regime. If, for example, an unexpected pregnancy results in a birth at a young age in 1968, a woman may choose and succeed at reducing the probability of a pregnancy in subsequent years, and still achieve the same lifetime count of children. For a discussion of the theoretically ambiguous impact of abortion restrictions on

birth rates, see Lawson and Spears (2025). Of course, the extent of persistence from period fertility to completed fertility depends on the details: A shock that encourages earlier-than-desired births, as Romania's might have, allows for adjustment later in life. But it may be harder, later in life, to adjust for a policy or event shock that leads to fewer births early in life.

Panel B of Figure 7 plots completed cohort fertility. As in earlier figures, cohorts are plotted along the horizontal axis according to the year in which they turned 30. Although Romanian completed cohort fertility began at a higher level than in Bulgaria over the available data series, completed cohort fertility in Romania did not maintain a sizable upward trend relative Bulgaria during the period that Decree 770 was in force. Importantly, there is no straightforward difference-in-differences specification that could be applied to the data in panel B: Even though we have arranged cohorts here to align with the year in which they turned 30, women who were, for example, 29 years-old at the date of the policy change would have been affected as well.

Over the period of the Decree 770 regime, cohort birth rates in Romania were never far from where they started. When the policy ended in 1989 with the ouster of Romania's dictator (coincident with the international fall of communism in Central and Eastern Europe), cohort and period fertility began a period of decline. Decree 770 had many obvious and harmful effects. And yet, it is difficult in these data to establish exactly what quantitative impact Decree 770 had on completed cohort fertility. What is clear is that analyzing this extreme policy in terms of period total fertility rates would be substantially misleading in terms of its effects on lifelong birth rates.

There is no simple solution to the cohort-period challenge, though a deeper engagement with the tools of population science may aid researchers in developing a collage of evidence. For example, a middle ground between examining period total fertility rates and completed cohort fertility is to consider "tempo-adjusted" period fertility rates, which attempt to account for how much of the movement in period fertility rates might be due to changes in timing of births (Bongaarts and Feeney 1998). As with any method of adjustment that relies on modeled assumptions, the quality of the assumption will influence the outcome. A related approach is that, rather than only quantifying effects on potential parents' average number of children, researchers can study parity progression (meaning progression by birth order): Do parents with one child have a second? Do parents with two children have a third? Kearney and Levine (2025) take this approach in a recent study of US fertility and find that, even though it is too early to know younger cohorts' completed birth rates, there is little age span left, on average, to achieve large family sizes.

Another research strategy uses survey questions that ask people how many children they want. To be sure, stated preferences must be interpreted with care. But evidence suggests that on average, populations do tend to have roughly the number of children that people in those populations report wanting (Pritchett 1994; Gietel-Basten et al. 2024). That certainly does not mean that individuals exercise perfect control over their fertility. But it is evidence that a rebound in birth rates

is unlikely in populations where, on average, young adults do not report wanting a replacement-level number of children.

Discussion

Birth rates have been falling. Not only for the past few years or decades, not only in a few countries, but worldwide and for as long as reliable data exist. There is no reason found in evidence to expect a certain, automatic reversal—not from the evidence in the discipline-spanning literature on long-run fertility trends and policy impacts, and not in the evidence presented in this article. To put it bluntly, history offers no examples of societies recognizing very low birth rates as a social priority and then responding with effective changes that restore, and sustain, replacement-level fertility. If changes that would bring global birth rates back to replacement after they fall below it are unlikely, then a long period of global depopulation would be a likely future. Our times, in which a large number of people exist on the planet, would be a historical anomaly that will come to an end.

These observations do not rule out the possibility of a rebound. Nor do they imply that researchers have hit a dead end, so nothing more can be learned about fertility decline or the potential for reversals. The problem of persistently low global fertility is new at a historical scale. From a scholarly perspective, it is understudied. From a practical perspective, no countervailing policies or social movements have arisen commensurate with the scale of this emergent challenge. It should not be surprising that packages of policies that amount to a few thousand dollars do not stack up as decisive in childbearing decisions when weighed against the life-defining choices about what sort of family to build.

What about much more transformative changes, larger than anything yet seen, that would make parenting more supported and valued, fairer, and more readily compatible with other aspirations for life—education, career, other relationships, and life projects? Could that avoid global depopulation? The econometric evidence cannot address this question because only small efforts have been tried, and this question is not about small efforts. Elsewhere, we have written: "Some policies have tweaked the edges of our societies and economies. But nobody has tried anything that would adequately challenge conventions, challenge social orders, and challenge what gets society's attention, power, and investment. No revolution has yet envisioned a future in which everyone who benefits from a healthy, joyful childhood looks forward to sharing in the work of giving one to the next generation" (Spears and Geruso 2025). We hope such a change someday occurs. But choosing that sort of investment would involve a massive reorganization of social priorities—beyond an extension of parental leave, beyond a modest child tax credit, beyond subsidized childcare, and so on. There is no reason to assume that societies will choose to do so much, even if birth rates continue to hit new lows.

The possibility of a future in which the global population shrinks—generation by generation, with no definite end—suggests a further evaluative question which

this paper has not addressed: Would long-term global depopulation be a future to welcome, or would it be better if the population stabilized, instead (perhaps at some smaller size than today's)? Although this question is beyond our focus in this essay, we have considered it in depth in *After the Spike* (Spears and Geruso 2025), where we conclude that stabilization, someday, would be a better future from a social welfare perspective, rather than ongoing depopulation. Elsewhere in this symposium, David Weil makes a somewhat contrasting case, focusing on implications for US per capita incomes. We emphasize different consequences of depopulation than Weil does. One reason is that our focus is global and on a longer-term future. Another reason is that the social welfare evaluation we endorse assigns positive weight to both the quality and quantity of lives lived. We encourage readers to consider this long-term future, and—whatever their normative conclusions—to look ahead to the likely effects and consequences of an enduring, unprecedented change from global population growth to indefinite global population decline.

■ *This research was supported by grant P2CHD042849 awarded to the Population Research Center at UT Austin by the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the views of the National Institutes of Health. The relevant section of this article subsumes the working paper "Childless or Child-fewer? Childlessness and Parity Progression where Fertility is Below Replacement" (Geruso and Spears 2025). We are grateful for comments from Diane Coffey and John Ermisch.*

References

Aaronson, Daniel, Fabian Lange, and Bhashkar Mazumder. 2014. "Fertility Transitions along the Extensive and Intensive Margins." *American Economic Review* 104 (11): 3701–24.

Albanesi, Stefania, and Claudia Olivetti. 2014. "Maternal Health and the Baby Boom." *Quantitative Economics* 5 (2): 225–69.

Bailey, Martha J., and William J. Collins. 2011. "Did Improvements in Household Technology Cause the Baby Boom? Evidence from Electrification, Appliance Diffusion, and the Amish." *American Economic Journal: Macroeconomics* 3 (2): 189–217.

Beaujouan, Éva, Kryštof Zeman, and Mathías Nathan. 2023. "Delayed First Births and Completed Fertility across the 1940–1969 Birth Cohorts." *Demographic Research* 48: 387–420.

Bongaarts, John, and Griffith Feeney. 1998. "On the Quantum and Tempo of Fertility." *Population and Development Review* 24 (2): 271–91.

Brown, Andrea D., Brady E. Hamilton, Dmitry M. Kissin, and Joyce A. Martin. 2025. "Trends in Mean Age of Mothers: United States, 2016–2023." *National Vital Statistics Reports* 74 (9): 1–6.

Coale, Ansley Johnson. 1986. "The Decline of Fertility in Europe since the Eighteenth Century as a Chapter in Human Demographic History." In *The Decline of Fertility in Europe*, edited by Ansley Johnson Coale and Susan Cotts Watkins, 1–30. Princeton University Press.

Deaton, Angus. 2013. *The Great Escape: Health, Wealth, and the Origins of Inequality*. Princeton University Press.

Dettling, Lisa J., and Melissa Schettini Kearney. 2025. "Did the Modern Mortgage Set the Stage for the US Baby Boom?" NBER Working Paper 33446.

Doepke, Matthias, Anne Hannusch, Fabian Kindermann, and Michèle Tertilt. 2023. "The Economics of Fertility: A New Era." In *Handbook of the Economics of the Family*, Vol. 1, edited by Shelly Lundberg and Alessandra Voena, 151–254. North-Holland.

Doepke, Matthias, Moshe Hazan, and Yishay D. Maoz. 2015. "The Baby Boom and World War II: A Macroeconomic Analysis." *Review of Economic Studies* 82 (3): 1031–73.

Ehrlich, Paul R. 1968. *The Population Bomb*. Sierra Club and Ballantine Books.

Ermisch, John. 2025. "The Role of Childlessness in Changes in English Cohort Fertility." Unpublished.

Espenshade, Thomas J., Juan Carlos Guzman, and Charles F. Westoff. 2003. "The Surprising Global Variation in Replacement Fertility." *Population Research and Policy Review* 22 (5): 575–83.

Fogel, Robert William. 2004. *The Escape from Hunger and Premature Death, 1700–2100: Europe, America, and the Third World*. Cambridge University Press.

Gauthier, Anne H., and Stuart Gietel-Basten. 2025. "Family Policies in Low Fertility Countries: Evidence and Reflections." *Population and Development Review* 51 (1): 125–61.

Geruso, Michael, and Dean Spears. 2025. "Childless or Child-fewer? Childlessness and Parity Progression Where Fertility Is Below Replacement." NBER Working Paper 33913.

Geruso, Michael, and Dean Spears. 2026. *Data and Code for: "The Likelihood of Persistently Low Global Fertility."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E239496V1.

Gheţău, Vasile. 1997. "Evoluţia fertilităţii în România: De la transversal la longitudinal." *Revista de Cercetări Sociale* 1: 3–85.

Gietel-Basten, Stuart. 2019. *The "Population Problem" in Pacific Asia*. Oxford University Press.

Gietel-Basten, Stuart, Melissa LoPalo, Dean Spears, and Sangita Vyas. 2024. "Do Fertility Preferences in Early Adulthood Predict Later Average Fertility Outcomes of the Same Cohort?: Pritchett (1994) Revisited with Cohort Data." *Economics Letters* 244: 111975.

Gietel-Basten, Stuart, and Sergei Scherbov. 2019. "Is Half the World's Population Really Below 'Replacement-Rate'?" *PLoS ONE* 14 (12): e0224985.

Gietel-Basten, Stuart, Dean Spears, and Leela Visaria. 2022. "Low Fertility with Low Female Labor Force Participation in South India." Paper presented at the Population Association of America 2022 Annual Meeting, Atlanta, April 8.

Hajnal, John. 1947. "The Analysis of Birth Statistics in the Light of the Recent International Recovery of the Birth-Rate." *Population Studies* 1 (2): 137–64.

Hellstrand, Julia, Jessica Nisén, Vitor Miranda, Peter Fallesen, Lars Dommermuth, and Mikko Myrskylä. 2021. "Not Just Later, but Fewer: Novel Trends in Cohort Fertility in the Nordic Countries." *Demography* 58 (4): 1373–99.

Hellstrand, Julia, Jessica Nisén, and Mikko Myrskylä. 2022. "Less Partnering, Less Children, or Both? Analysis of the Drivers of First Birth Decline in Finland since 2010." *European Journal of Population* 38 (2): 191–221.

Human Fertility Database. 2023. *Human Fertility Database*. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). https://www.humanfertility.org (accessed February 2025).

Hwang, Jisoo. 2023. "Later, Fewer, None? Recent Trends in Cohort Fertility in South Korea." *Demography* 60 (2): 563–82.

IIPS (International Institute for Population Sciences) and ICF. 2021. *India: Standard DHS, 2019–21* (*NFHS-5*). DHS Program, ICF. https://dhsprogram.com/methodology/survey/survey-display-541.cfm (accessed March 2025).

International Monetary Fund. 2025. *World Economic Outlook: Global Economy in Flux, Prospects Remain Dim*. IMF.

Kaneda, Toshiko, and Carl Haub. 2022. "How Many People Have Ever Lived on Earth?" Population Reference Bureau, November 15. https://www.prb.org/articles/how-many-people-have-ever-lived-on-earth/.

Kearney, Melissa Schettini, and Phillip B. Levine. 2025. "Why Is Fertility So Low in High Income Countries?" NBER Working Paper 33989.

Kearney, Melissa S., Phillip B. Levine, and Luke Pardue. 2022. "The Puzzle of Falling US Birth Rates

since the Great Recession." *Journal of Economic Perspectives* 36 (1): 151–76.
Kirmeyer, Sharon E., and Brady E. Hamilton. 2011. "Childbearing Differences among Three Generations of US Women." NCHS Data Brief 68.
Kitagawa, Evelyn M. 1955. "Components of a Difference between Two Rates." *Journal of the American Statistical Association* 50 (272): 1168–94.
Lawson, Nicholas, and Dean Spears. 2025. "Equilibrium Effects of Abortion Restrictions on Cohort Fertility: Why Restricting Abortion Access Can Reduce Human Capital, Social Welfare, and Lifetime Fertility Rates." *Journal of Economic Behavior and Organization* 238: 107216.
Lesthaeghe, R. 1995. "The Second Demographic Transition in Western Countries: An Interpretation." In *Gender and Family Change in Industrialized Countries*, edited by Karen Oppenheim Mason and An-Magritt Jensen, 17–62. Oxford University Press.
Lewis-Kraus, Gideon. 2025. "The End of Children." *New Yorker*, February 24. https://www.newyorker.com/magazine/2025/03/03/the-population-implosion.
Mathews, T. J., and Brady E. Hamilton. 2002. "Mean Age of Mother, 1970–2000." *National Vital Statistics Reports* 51 (1): 1–16.
Mathieu, Edouard, and Lucas Rodés-Guirao. 2022. "What Are the Sources for Our World in Data's Population Estimates?" Our World in Data. https://ourworldindata.org/population-sources.
Notestein, Frank W. 1945. "Population—The Long View." In *Food for the World*, edited by Theodore W. Schultz, 36–77. University of Chicago Press.
Olivetti, Claudia, and Barbara Petrongolo. 2017. "The Economic Consequences of Family Policies: Lessons from a Century of Legislation in High-Income Countries." *Journal of Economic Perspectives* 31 (1): 205–30.
Park, Narae, Sangita Vyas, Kathleen Broussard, and Dean Spears. 2023. "Near-Universal Marriage, Early Childbearing, and Low Fertility: India's Alternative Fertility Transition." *Demographic Research* 48 (34): 945–56.
Pritchett, Lant H. 1994. "Desired Fertility and the Impact of Population Policies." *Population and Development Review* 20 (1): 1–55.
Sobotka, Tomáš. 2021. "World's Highest Childlessness Levels in East Asia." *Population and Societies* 595: 1–4.
Sobotka, Tomáš, and Éva Beaujouan. 2017. "Late Motherhood in Low-Fertility Countries: Reproductive Intentions, Trends and Consequences." In *Preventing Age Related Fertility Loss*, 11–29. Springer.
Sobotka, Tomáš, Anna Matysiak, and Zuzanna Brzozowska. 2019. "Policy Responses to Low Fertility: How Effective Are They?" UNFPA Technical Division Working Paper 1.
Spears, Dean, and Michael Geruso. 2025. *After the Spike: The Risks of Global Depopulation and the Case for People*. Simon and Schuster.
Spears, Dean, Sangita Vyas, Gage Weston, and Michael Geruso. 2024. "Long-Term Population Projections: Scenarios of Low or Rebounding Fertility." *PLoS ONE* 19 (4): e0298190.
Stone, Lyman. 2020. "The Rise of Childless America." Institute for Family Studies (blog), June 4. https://ifstudies.org/blog/the-rise-of-childless-america.
United Nations. 2024. *World Population Prospects 2024*. United Nations Department for Economic and Social Affairs, Population Division. https://population.un.org/wpp/ (accessed July 2025).
United Nations, Department of Economic and Social Affairs, Population Division. 2018. World Population Policies 2015 (ST/ESA/SER.A/374): p. 84.
US Department of Health, Education, and Welfare. 1973. *Fertility Tables for Birth Cohorts by Color: United States, 1917–73*. DHEW Publication (HRA) 76-1152.
Van Bavel, Jan, and David S. Reher. 2013. "The Baby Boom and Its Causes: What We Know and What We Need to Know." *Population and Development Review* 39 (2): 257–88.
Visaria, Leela. 2022. "India's Date with Second Demographic Transition." *China Population and Development Studies* 6 (3): 316–37.
Weston, Gage. 2024. *PWI-Population-Projections (commit 4b31f25)*. Github repository. https://github.com/gagemweston/PWI-Population-Projections/ (accessed July 20, 2025).
Wolfe, Rachel. 2024. "Why Americans Aren't Having Babies." *Wall Street Journal*, July 20. https://www.wsj.com/lifestyle/relationships/americans-babies-childless-birthrate.
Zeman, Kryštof, Éva Beaujouan, Zuzanna Brzozowska, and Tomáš Sobotka. 2018. "Cohort Fertility Decline in Low-Fertility Countries: Decomposition Using Parity Progression Ratios." *Demographic Research* 38: 651–90.

Journal of Economic Perspectives—Volume 40, Number 1—Winter 2026—Pages 27–46

How Much Would Continued Low Fertility Affect the US Standard of Living?

David N. Weil

In 2023, the total fertility rate (TFR) in the United States was 1.62 births per woman (Osterman et al. 2025). That figure was not only the lowest in the country's history, but also well below the replacement rate of 2.08 births per woman that would be required to maintain constant population size in the long run with current mortality rates and absent any net migration. The United States previously experienced a bout of sub-replacement fertility between 1972 and 1989, but that dip was not severe and came after the post–World War II "Baby Boom," during which the TFR had been well above three. The current episode of low fertility, which started with a trend break at the time of the Great Recession, is expected by many forecasters to persist into the foreseeable future. This expectation is based largely on the ubiquity and durability of sub-replacement fertility among developed countries worldwide—a situation in which the United States had long been the most notable outlier (Kearney, Levine, and Pardue 2022; Geruso and Spears in this symposium).

Low fertility, both in the United States and around the world, is frequently discussed as a crisis. Commentators worry that the burden of supporting the dependent elderly will crush the economic hopes of working age adults, that smaller populations will be unable to service the debt accumulated by earlier generations, and that innovation will dry up. Some of the concerns transcend economics. Pope Francis (2022) stated that "this denial of fatherhood or motherhood diminishes us, it takes away our humanity. And in this way civilization becomes aged and without

■ *David N. Weil is James and Merryl Tisch Professor of Economics, Brown University, Providence, Rhode Island. His email address is David_Weil@Brown.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251462.

humanity, because it loses the richness of fatherhood and motherhood. And our homeland suffers, as it does not have children."

In this paper, I focus primarily on the question of how sub-replacement fertility affects the standard of living, more specifically the level of age-adjusted consumption per capita. Because sub-replacement fertility is such a recent phenomenon, and because its effects take a very long time to manifest in economic outcomes, my primary approach is to use off-the-shelf economic and demographic models to study its impact. I start by examining the effect of fertility on the relative numbers of people in different age groups as well as the population growth rate, both in demographic steady states and along transition paths following a change in fertility. I then consider how these variations in population age structure and growth rate affect the standard of living through channels of transfers among different age groups, effects on investment, and the servicing of government debt. I also examine economic effects that are functions of the size of the population, rather than its age structure or growth rate. These are specifically the interaction of population with the environment and the speed of technological progress. Finally, I discuss the extent to which raising fertility would help repair the financing of Social Security.

My conclusion is that much of the worry about sub-replacement fertility is overstated. Quantitatively, the net effect of even a large fertility reduction on the US economy would be a relatively small decline in the standard of living. Comparing demographic steady states and focusing on the most easily quantified channels, a version of the United States with a total fertility rate of one child per woman would have consumption per capita that was 8.7 percent lower than a version of the country where the TFR was two. In the first four decades of the transition following a decline in fertility below the replacement rate, consumption is actually higher than it would have been if fertility had remained constant. Indeed, much of the *sturm und drang* regarding the economic effects of current population aging is related to the ending of such a transitory period of good times that resulted from fertility declining from its Baby-Boom highs to near replacement, starting in the 1960s. Finally, it is important to note that any attempt to fix the economic problems stemming from low fertility by raising the birth rate will entail a period of higher overall dependency in the decades that it takes the resulting children to become productive adults.

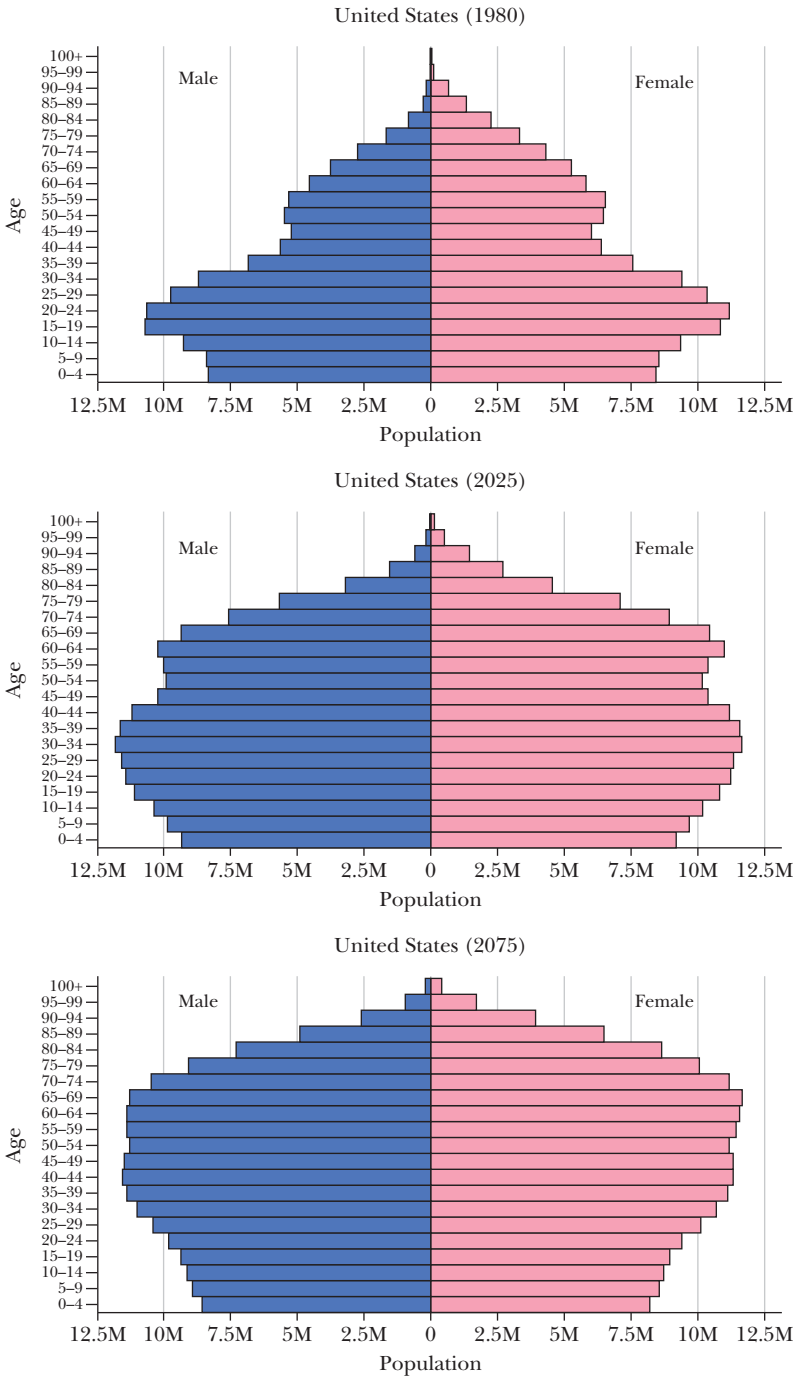
Fertility, Population Growth, and Consumption in Stable Populations

Figure 1 shows population pyramids for the United States for the years 1980, 2025, and projected for 2075. In 1980, the US population was 227.2 million and grew at a rate of 1.0 percent per year. In 2025, the population was 338.0 million, growing at 0.4 percent per year.

The projection for 2075, which assumes that the total fertility rate will remain at approximately 1.6 children per woman for the next 50 years, implies a total population of 369.0 million, growing at a rate of almost exactly zero.

Figure 1

Population Pyramids for the United States



Source: National Center for Health Statistics (2025), US Census Bureau (2025), and Social Security Administration (2022). Details of the calculations are in Weil (2026).

Note: The bars to the left show the male population, while bars to the right show the female population.

The slowing rate of population growth as well as the redistribution of population into older age groups that is evident in this figure are largely the result of declining fertility, but not entirely so. Falling mortality and international migration also play roles. The projection in Figure 1 assumes that life expectancy at birth will rise by 5.6 years and net international migration will range between 850,000 and one million people per year over the next half century (US Census Bureau 2023). Further, changes in fertility affect the country's demography with an extremely long lag. This lag is due in part to the slow pace at which humans move through their life cycles, but also results from the so-called "demographic momentum" that is built into the relative sizes of different birth cohorts. For example, the fact that current cohorts of young girls are smaller than those of women in their child-bearing years implies that the number of births will decline over time if age-specific birthrates remain constant. In what follows, I use several different approaches to reveal the effect of the current bout of low fertility on demographic and economic outcomes, separating this from the other factors just mentioned.

As a first step in isolating the effect of fertility on a country's demography, and thus on its economy, I apply a tool from demography, the *stable population*. A stable population is a theoretical distribution of population shares across ages that would eventually result if both age-specific fertility and age-specific mortality rates remained constant, and there was no migration. It can be thought of as a steady-state age distribution, with a corresponding steady-state rate of growth or shrinkage. To be clear, a stable population is *not* the same as a constant-sized population. Only if fertility is at the replacement rate will a stable population be of constant size.

Comparing stable populations with different fertility rates—assuming the same pattern of mortality—allows for a *ceteris paribus* comparison between different fertility scenarios, abstracting from the short run dynamics as observed in actual populations. Below I return to consider these short-run dynamics explicitly.

Table 1 shows data for a set of stable populations constructed using age- and sex-specific mortality rates for the United States for 2019, with total fertility rates ranging from one to three children per woman. For each TFR, the table shows the associated stable-population growth rates of population and the shares of the population in different age groups. One can think of these stable populations as representing a menu of different possible steady-state age structures.

I use these stable populations to illustrate and quantify three channels through which fertility affects the standard of living: age structure effects on transfers and the effect of population growth on both required capital investment as well as servicing government debt. I then discuss the dynamics of adjustment when fertility changes. Finally, I consider the economic effects of additional channels by which population size (rather than its age structure or growth rate) affect economic wellbeing.

Age Structure, Dependency, and Transfers

The age structure of the population—that is, the relative sizes of different age groups—affects the standard of living because of the nature of the human life cycle: In advanced economies, children and old people on average consume more than

Table 1
Stable Populations for the United States

TFR	N (%)	Share 0–19 (%)	Share 20–64 (%)	Share 65+ (%)	Adj. support ratio
1.00	−2.4	10.0	47.9	42.2	0.490
1.25	−1.7	13.6	51.0	35.4	0.526
1.50	−1.1	17.3	52.8	30.0	0.549
1.75	−0.6	20.8	53.6	25.6	0.563
2.0	−0.1	24.1	53.9	22.1	0.570
2.25	0.3	27.2	53.7	19.2	0.571
2.50	0.6	30.1	53.2	16.8	0.570
2.75	0.9	32.8	52.5	14.8	0.565
3.00	1.2	35.3	51.6	13.1	0.559

Source: Details of the calculations are in Weil (2026).
Note: Stable populations are constructed using US age- and sex-specific mortality from 2019 and proportionally scaling age-specific fertility from 2019.

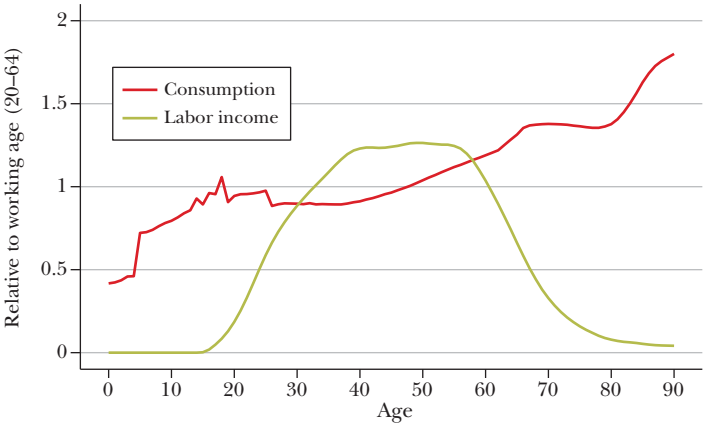
they produce, with people in the middle of their lives doing the opposite. Shifting population between these age groups thus impacts consumption possibilities.

To illustrate this point, consider a very simple economic-demographic model. Divide the population into three groups: young (aged 0–19), working-age (20–64), and old (65+). Assume that all people of working age do in fact work, that output is produced solely with the labor of working age people, that people of all ages have equal consumption, and that total consumption and total output are equal in each period. Finally, take the level of income per working age person as exogenous and unrelated to the age structure of the population. In this simple setting, consumption per capita is just equal to output per worker multiplied by the share of the population that is of working age.

Table 1 shows the fundamental tradeoff between fertility and the sizes of different dependent age groups. When fertility is high and population is growing quickly, the fraction of the population that is made up of elderly people is small, but the fraction made up of children is large. When fertility is low, the situation is reversed. From looking at Table 1, it is clear that the fertility rate that minimizes the total dependency burden—that is, the one that maximizes the working-age share of the population—is in fact quite close to the replacement rate. To be clear, this is not a generic result; it just happens to hold true for the particular mortality rate and age cutoffs between groups that were used in this exercise.

Of course, this very simple model makes strong assumptions: all working-age people produce the same amount, no one in other age groups produces anything, and everyone has the same level of consumption. To make the model more realistic, I use data from the National Transfer Accounts project (discussed in Lee et al. 2014). Figure 2 shows the average consumption and labor income by age for the United States in 2011. Consumption includes private expenditures as well as government services imputed by age; labor income includes wages as well as imputed labor income of the self-employed and unpaid family workers. Not surprisingly, Figure 2 shows that average consumption among old people is higher than

Figure 2
Age Profiles of Consumption and Labor Income



Source: National Transfer Accounts (2025).
Note: Consumption includes private expenditures as well as government services imputed by age; labor income includes wages as well as imputed labor income of the self-employed and unpaid family workers.

among working age, which in turn is higher than among children. This will tend to raise the economic burden of population aging that results from low fertility.

Using the data in Figure 2 along with data and projections of the number of people at each age by year, I construct a measure of the *adjusted support ratio*. This is the weighted sum of population by age group, where the weights are average labor income by age, divided by a similar sum where the weights are average consumption by age. Both labor income and consumption are normalized relative to the average for working-age adults. Conceptually, the adjusted support ratio functions just as the working-age share of the population in the simple calculations above: The larger its value, the higher will consumption will be relative to the earnings of working-age adults. Weil (1997) goes through the algebra of this approach in detail.

The last column in Table 1 shows adjusted support ratios in stable populations. The total fertility rate that maximizes the adjusted support ratio is 2.24. This finding suggests that from the perspective of maximizing age-adjusted consumption, fertility below the replacement level is indeed a bad thing.

In addition, Table 1 is also useful for showing the magnitude of the dependency cost of being away from the optimum. To pursue this question, I will focus on a somewhat stark comparison: a total fertility rate of two versus one. The former, as the table shows, is only slightly below the level of fertility that maximizes adjusted consumption. It is also, of course, just below the replacement rate. The latter is by most accounts an extremely low level—far below the current TFR in the United States, although still above South Korea, which in 2022 had a TFR of 0.78. Calculating the magnitude of the effect of going from a TFR of two to a TFR of one just involves looking at the adjusted support ratios in the two cases. The relevant values are .570 for a TFR of two

and .490 for a TFR of one, implying that consumption in the low-fertility scenario is 14.0 percent lower than when the TFR is just below replacement.

Before going further, it is worth noting that because of the concavity of the adjusted support ratio, deviations from optimal fertility that are smaller than the one just analyzed have much less significant effects. For example, consumption in a scenario with a total fertility rate of 1.5 is only 4 percent lower than when TFR is two. Later in the paper, I will also consider some economic effects of demography beyond population age structure that will somewhat alter the calculus of what TFR maximizes consumption.

Investment and Capital Accumulation

In 2023, gross domestic investment in the United States (inclusive of government investment) came to 21.5 percent of GDP (Council of Economic Advisers 2025, Table B19). Investment can be conceptually divided into three categories: replacing depreciated capital, increasing the capital/labor ratio, and, if the labor force is growing, supplying new workers with capital. If the labor force grows more slowly, and even more so if it shrinks, the amount of investment in this last category can be reduced, which expands the consumption possibilities of the economy.

In general, tracing the effect of slower population growth on investment is complex. In a neoclassical growth model in which the social planner's welfare function is defined over average utility from consumption, slowing population growth induces a rise in the capital/labor ratio, as the investment cost of maintaining this higher ratio is reduced. Cutler et al. (1990) show that this effect goes away if the social planner cares about total rather than average welfare—an issue to which I return below. In an economy with decentralized saving and investment decisions, the impact of demographic change on investment will depend on, among other things, the extent to which the capital market is open to the world and the model determining individual saving.

Rather than delve into these complexities, I pursue a quantitative shortcut that captures the essence of the issue. First, I assume that the growth rates of the total population and the labor force are equal, as holds in the case of a stable population. Consider a path for investment expenditures relative to GDP (I/Y) that, for a given growth rate of the population, would yield a particular path of the capital/output ratio (K/Y). If we then ask what would happen to required I/Y in order to maintain that same path of K/Y if there were a different growth rate of population, the answer is simple: I/Y would have to change by an amount equal to the desired K/Y ratio multiplied by the change in population growth.

To give a concrete example, consider the comparison of a stable population with a total fertility rate of one versus a stable population with a TFR of two. From Table 1, the former has a growth rate of total population that is 2.3 percent per year lower than the latter. The K/Y ratio in the United States in 2019 was approximately 3.3 according to the Penn World Tables (Version 10.01, updated from Feenstra, Inklaar, and Timmer 2015). Thus, the version of the United States with a TFR of one would have investment required to maintain capital per worker that would be

$3.3 \times 2.3\% = 7.6\%$ GDP lower than the version of the United States with TFR of two. This is a large effect.

In addition to the change in required investment, demographic change due to lower fertility may have other related macroeconomic impacts that are of interest. One that has received attention in the literature is a reduction in the real interest rate, via both decreased investment demand and the shifting of population into ages with higher average wealth holdings (Eggertsson, Mehrotra, and Robbins 2019). Tracing the welfare effect of such a reduction is difficult, but it is certainly the case that lower interest rates do not qualify as being bad for the economy in a self-evident sense.

Servicing Government Debt

My analysis of debt sustainability parallels that of capital accumulation, although the direction of the effect is reversed. In an economy with a growing population, the labor force will be larger over time and GDP will rise as a result. Thus, the debt/GDP ratio will decline over time; more concretely, an economy in this population growth scenario is required to run a smaller primary surplus (or may run a larger primary deficit) in order to achieve a given path of the debt/GDP ratio. When population growth becomes negative, this process goes into reverse: the size of the labor force declines, GDP growth slows, and the primary surplus has to be increased to maintain a given debt/GDP ratio.

Of course, in the current context of US fiscal policy, it may seem fanciful to discuss stabilization of the debt/GDP ratio, but the same logic regarding the role of population growth holds even when considering paths of debt/GDP that are not flat. Consider a path of the primary budget surplus relative to GDP that, for a given growth rate of the population and labor force, would yield a particular path of the debt/GDP ratio. If we then ask what would happen to the primary surplus that would be required to achieve this path with some different population growth rate, the answer is that the primary surplus would have to grow by the debt/GDP ratio multiplied by the *negative* of the change in the population growth rate.

Applying the same comparison of a version of the United States with a total fertility rate of one versus a version of the country with a TFR of two that was discussed above, we would multiply the difference in population growth rates (2.3 percent) by the debt/GDP ratio. In the United States, that ratio (specifically using debt held by the public) is almost exactly 1.0, and so the math is straightforward: in the low-TFR scenario, the primary surplus would have to be 2.3 percent of GDP larger than in the high fertility scenario in order to achieve the same path of the debt/GDP ratio.

Combining the Channels

Here I combine the three channels just discussed to produce a more comprehensive assessment of the economic impact of low fertility. For simplicity, I will continue to examine the case of stable populations; more concretely, I will compare outcomes of fertility rate of two and one.

The combined effect of these three channels is constructed by simply adding their individual effects. Comparing total fertility rates of one and two, the dependency

effect reduces consumption by 14.0 percent, the investment effect raises consumption by 7.6 percent, and the debt sustainability effect reduces consumption by 2.3 percent (with these latter two effects being scaled to the current K/Y and debt/GDP ratios of the United States). The combined effect of reducing the total US fertility rate from two to one is to reduce consumption by 8.7 percent.

An evolution of per capita consumption over a period of many decades such that it falls 8.7 percent below what it would otherwise have been is clearly unfortunate, but it falls short of being a crisis as is sometimes depicted. Indeed, the effect strikes me as relatively small. First, note that the fertility rates that I am comparing differ by an enormous amount. They were chosen for illustrative purposes rather than realism. For example, in the probabilistic projections of the United Nations (2024), a total fertility rate of 1.0 in the United States in the year 2100 is at the very bottom of the 95 percent probability interval, while a total fertility rate of 2.0 is at the top of the 80 percent probability interval. Even taking this extreme comparison, notice that 8.7 percent of GDP is slightly smaller than five-year's growth of US personal consumption per capita during the twenty-first century.[1] It is significantly smaller than the difference between consumption per capita in Canada and the United States.

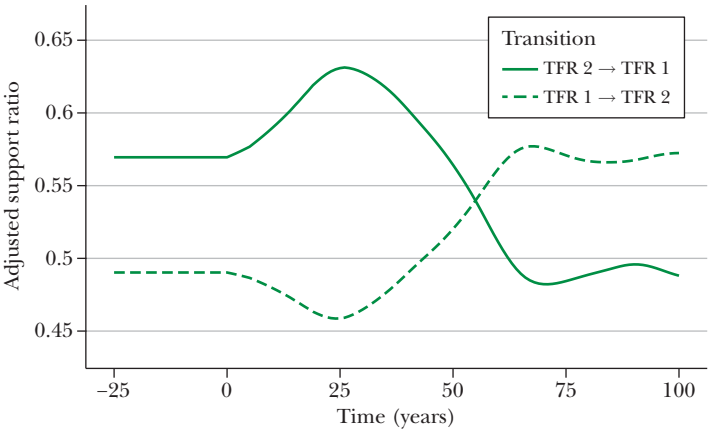
Transitional Dynamics

The analysis in the previous section focused on stable populations, that is, demographic steady states with constant fertility and mortality rates and no migration. In a stable population, the fractions of the population in different age groups are constant and the total population grows or shrinks at a constant rate. In such a setting, a change in fertility will eventually move to a new stable population, with a different age structure and a different growth rate. But this transition can take a long time and proceeds in a non-monotonic fashion. For example, when the fertility rate falls, the initial effect is to lower the ratio of children to working-age adults, without affecting the burden of old-age dependency. Only many decades later will the ratio of elderly to working age people rise.

To see why these transitional dynamics are important, we can go back to the simple economic-demographic model that was introduced earlier, in which the population is divided into three age groups with equal consumption and only the people in the middle group worked. In that setting, consumption per capita is equal to the product of output per worker and the working age fraction of the population. Further, the growth rate of consumption per capita is just the sum of the growth rate of output per worker and the growth rate of the working-age share of the population. A rising working-age share acts like a tailwind allowing consumption per capita

[1] Using data from the FRED database maintained by the Federal Reserve Bank of St. Louis, real personal consumption expenditures per capita over the period Q3 1990 to Q3 2019 grew at an annual rate of 1.76 percent.

Figure 3
The Demographic Dividend and the Demographic Debit



Source: National Center for Health Statistics (2025), US Census Bureau (2025), and Social Security Administration (2022). Details of the calculations are in Weil (2026).
Note: The graphs show the effect of an immediate jump upward or downward in Total Fertility Rate, starting with a stable population. Simulations are constructed using US age- and sex-specific mortality from 2019 and proportionally scaling age-specific fertility from 2019.

to grow more rapidly than income per worker while a falling working age share acts like a headwind, slowing consumption growth. Weil (1997) shows that this same effect holds in a richer model, in which both consumption and labor income vary with age as in Figure 2. In this case, the relevant tailwind or headwind is the growth rate of the adjusted support ratio.

Figure 3 illustrates these non-monotonic transitional dynamics in a demographic simulation, based on age-specific mortality rates from the United States in 2019 and assuming zero migration. The solid line shows the adjusted support ratio in a country that starts with a stable population and a total fertility rate of two. Beginning in year zero, the TFR jumps to one. As the figure shows (and consistently with Table 1), the long-run value of the support ratio at the new TFR is lower than at the initial value. However, the figure also shows a long period in which the adjusted support ratio is *higher* than its initial value. Bloom, Canning, and Sevilla (2003) call this transitory period of low dependency that arises following a decline in fertility the *demographic dividend*. The adjusted support ratio during the demographic dividend can exceed any ratio that is achievable in a stable population, but this happy period inevitably comes to an end. The decline in the adjusted support ratio that is currently taking place in the United States is in large part simply the fading away of the US demographic dividend, which was inevitable (Weil 1997).

This analysis of the dynamics of the support ratio in response to fertility changes has an unfortunate implication for pro-natalist policies. This is shown in the dashed line in Figure 3, which traces out the dynamics following a rise in the total fertility

rate from one to two. In this case, there is a long run increase in the adjusted support ratio, but a temporary decline. More specifically, it takes 41 years from the onset of the fertility increase for the adjusted support ratio to rise above its initial value. This phenomenon does not yet have a recognized name, but I propose calling it the *demographic debit*.[2] The existence of the demographic debit suggests that policies designed to raise fertility would not be appealing (at least on economic grounds) to elderly or near elderly voters, and that such policies would similarly be an unappetizing choice for politicians whose planning horizon rarely extends so far into the future.

Projected Effects of Low Fertility

As an alternative to the simulations used in the previous section, another way to examine the role of changing fertility is by varying birthrates in a projection model that starts with country's actual population age structure and allows for both ongoing migration and changing mortality. Figure 4 shows the adjusted support ratio in the United States through 2024 and then two different projections from the United Nations (2024). In the median projection the total fertility rate fluctuates between 1.6 and 1.65 for the remainder of the century. The "instant replacement" scenario, as the name implies, has the TFR jump immediately to the replacement level.

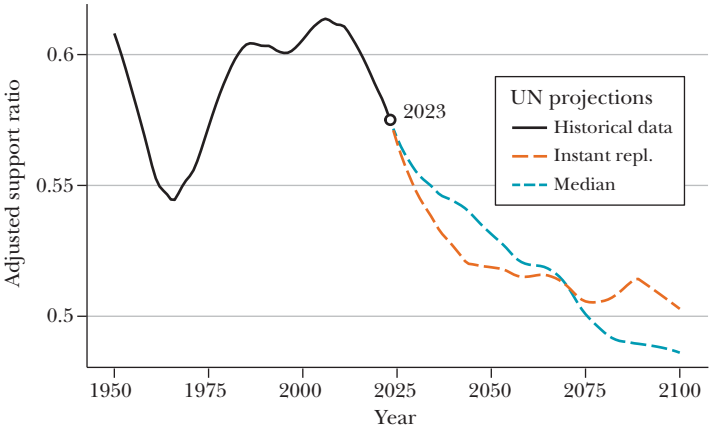
The first thing to note about Figure 4 is the pattern of variation in the adjusted support ratio that has taken place in the last several decades and that is projected for the next several. Although the picture is not as clean as what emerges from the simulation model in Figure 3, the message is very similar. Following the large reduction in fertility rates in the 1960s, there was a protracted rise in the adjusted support ratio that allowed for the growth of consumption per capita to exceed the growth of GDP per worker. This is exactly the demographic dividend of Bloom, Canning, and Sevilla (2003). Further, just as the model predicted, the demographic dividend eventually came to an end.

The adjusted support ratio reached a local maximum in 2006. Over the period 1965–1985, it rose at a particularly rapid rate of 0.52 percent per year, allowing for rapid growth of consumption per capita. Unfortunately, this tailwind has now turned into a headwind. Over the period 2025–2075, the adjusted support ratio is projected to fall at a pace of 0.25 percent per year, meaning that consumption per capita (adjusted for age-specific needs) will grow one quarter of a percent per year more slowly than income per worker.

The second key finding in Figure 4 is that for many decades, the adjusted support ratio is significantly *lower* in the instant replacement scenario than in the scenario where fertility remains below replacement. The gap between the two scenarios is largest 20 years after the two scenarios diverge, reflecting the fact that up to that

[2] Kotschy and Bloom (2023) use the phrase *demographic drag* to describe the reduction in the growth of output per capita relative to output per worker that results when old-age dependency rises in response to low fertility several decades earlier.

Figure 4
Adjusted Support Ratio by UN Fertility Scenario



Source: United Nations (2025).
Note: The projections give population by broad age groups: 0–19, 20–64, and 65+. I apply the average labor income and consumption weights from Figure 2 to these age groups.

point the high-fertility scenario features more children, but the same numbers of working-age and old people as the low-fertility scenario. To put this conclusion another way, a pro-fertility policy that succeeded in accomplishing an immediate sharp rise in the birthrate would lead to a lower support ratio for several decades. Eventually, higher fertility leads to a higher adjusted support ratio by lowering the share of the population made up of old people, but in the Figure 4 scenario that does not happen until 2071. This is exactly the implication of the demographic debit effect that was introduced in the simulation model above.

Economic Consequences of Fewer People

Lower fertility implies, over time, a smaller population. Separately from the growth-rate and age-structure effects examined above, population size can also have a direct effect on economic well-being. In this section, I examine two such channels.

Will a Smaller Population Help Mitigate Environmental Concerns?

The interaction of population size with natural resources is the issue on which the field of population economics cut its teeth, starting with Malthus (1798). In a constant-returns production function with land or another fixed factor as an input, a smaller population size leads directly to a higher standard of living. This issue of the ratio of population to resources is what energized population pessimists during the second half of the twentieth century (for example, Ehrlich 1968; Hardin 1995). While these scholars focused particularly on developing countries, they were also

quite clear in their assessment that richer countries like the United States would have a higher standard of living if they had fewer people.

There were a number of reasons why the dire predictions of the population pessimists did not pan out: for example, the possibilities of substitution among resource inputs in production, technological progress that allowed for new productive techniques, and the ability of highly populated but resource-poor countries such as Japan to use trade as a substitute for locally available inputs. Further, an analysis of how countries' populations compare to their available resources suggests that the United States is far from "full." For example, the US ratio of people to land area, with the latter adjusted to reflect characteristics that support economic activity, is about 40 percent of the world average (Henderson, Storeygard, and Weil 2020).

To the extent that the issue of population size relative to available resources remains relevant today, it is in two contexts. First, many of the poorest countries in the world today are still largely autarkic, where costs of or other barriers to trade are high, most of the population is dependent on small-scale farming or herding, and land degradation from overuse is a serious problem. Given such constraints, these countries would benefit economically from a smaller population, although they almost all still have fertility rates that are at or above the replacement level. By contrast, almost all countries with fertility well below replacement level are open to trade, with only a small fraction of the population deriving its livelihood from land or other fixed resources.

The population/resource ratio is also relevant for the world as a whole, albeit in a different way. While population pessimists of the twentieth century worried about the ratio of people to farmland and subsoil resources, current analyses of the interaction of population and the natural environment often focus on the emissions of greenhouse gases as drivers of climate change. From a climate change perspective, smaller populations in highly developed countries, which have the highest levels of greenhouse gas emissions per capita, are a good thing. Whether changes in fertility from the present moment onward would have a significant impact on cumulative greenhouse gas emissions depends on the path of decarbonization going forward. Many projections have the property that greenhouse gas emissions per unit of output will fall very quickly over the next 75 years, thus avoiding the worst effects of climate change (Budolfson et al. 2025). In those projections, conceivable variations in fertility in the wealthiest countries going forward will have only a minor additional impact on cumulative emissions as of the end of the century. Of course, if rapid decarbonization in per-capita terms did not happen, then a declining population would indeed mitigate the resulting climate disaster.

Will the Fewer People Mean Slower Technological Progress?

In just about every model that economists construct, technological progress is the driver of economic growth over the very long run. New technologies have the property that they are non-rival, so that benefits from an invention are scaled by the population that has access to it. In turn, creation of new technologies requires resources in the form of labor, human capital, and physical capital. An economy with more people

will, all else being equal, have more people working on creating new technologies, and thus faster economic growth. Jones (2022) argues that for this reason, zero or negative population growth will have a negative impact on the long-run standard of living. This argument is a reasonable one, but several caveats suggest that it may not be of great quantitative importance, at least for horizons of less than a century.

A starting point for thinking about this issue is to note that the rate of technological progress relevant for the US standard of living going forward is shared among the group of rich, technologically advanced countries as a whole. Within this cutting-edge group, there are few barriers to technology transmission. While there are clearly localized spillovers from research and development, the long-run growth rate of total factor productivity that results from ideas emerging in Silicon Valley will be roughly the same in Beijing as in Boston. In 2021, the United States accounted for 19 percent of the world's science and technology researchers (OECD 2025, full time equivalents). Low fertility and a smaller population in the United States will thus have a relatively small effect on the world's research and development effort or the pace of technological progress for the world as a whole.

Of course, low fertility will also eventually reduce the size of the research and development workforce in the rest of the world as well, but this will not happen for a while, for two reasons. First, many new countries will be joining the group that is at the technological cutting edge and thus contributing to technological progress that is relevant for the United States. Second, there will be rising human capital of the labor force, and thus growth in the number of potential researchers, in populous countries such as China and India. I can put some quantitative flesh on this point by looking not at research and development workers directly (a group for which I do not know of any projections), but at the number of people with post-secondary education, which is the pool from which these workers will be drawn. The combined population of the regions where the vast bulk of current research and development takes place (Europe, North America, East Asia, Australia, and New Zealand) is projected to decline by 22 percent between 2050 and 2100, while the population of the world as a whole is projected to increase by only 3.1 percent However, over this same period, the number of people with post-secondary education worldwide is projected to increase by 78 percent, according to the Wittgenstein Centre for Demography and Global Human Capital (KC et al. 2024, Medium SSP2 scenario).

Finally, the Jones (2022) model implies that the transition to zero or negative population growth affects the speed of technological progress with an extremely long lag. Calculations in Budolfson et al. (2025) can be used to illustrate this point. They compare the speed of technological progress in two population scenarios. The "depopulation" scenario matches and extends the UN median projection, with global total fertility rates converging toward 1.66 children per woman. World population peaks in the mid-2080s around 10.3 billion and falls permanently thereafter. In the "stabilization" scenario, TFRs are bounded from below by the replacement rate, implying an immediate jump upward in all countries where they are currently below that level. By the year 2200, the world population is nearly twice as large in the stabilization scenario as in the depopulation scenario.

Feeding these different population scenarios through a Jones-style model of endogenous growth produces faster technological progress in the scenario where there are more people. But because of the slow speed with which fertility affects population growth, as well as the decline in research productivity as knowledge accumulates that is integral to the Jones model, the magnitude of the effect is not large. In the year 2100, total factor productivity is only 3 percent higher in the stabilization scenario than under depopulation, and in the year 2200 the difference is only 14 percent. This is a relatively small effect, given that even in the depopulation scenario, total factor productivity increases by a factor of six between now and 2200.

Can Raising Fertility Fix the Solvency Problem of Social Security?

The bulk of this article has examined the general economic impact of low fertility, focusing on the consumption needs of people of all age groups. Notably, from this perspective, the effect of low fertility in raising old-age dependency is balanced by both the reduction in the need to support children and the decline in required investment as labor-force growth slows.

For many commentators, however, the much more salient issue is how low fertility affects government support for the elderly—in particular, Social Security in the United States. This focus is not unreasonable: under present projections, the Old-Age and Survivors Insurance trust fund will be exhausted in 2033; in the absence of legislative changes, this will trigger an immediate 23 percent cut in benefits.

To illustrate the impact of low fertility, I take advantage of tables from the report of the actuaries of the Social Security Administration (2025), which calculate the 75-year actuarial balance of Social Security under a baseline set of assumptions and then under a series of scenarios that vary one or another assumption. Under baseline assumptions, the system's 75-year actuarial balance is −3.82 percent. This actuarial balance indicates the approximate amount that the payroll tax rate would have to immediately rise above its current level of 12.4 percent in order to keep the system solvent for the next three-quarters of a century (the exact amount of the required rise is slightly smaller, for complicated reasons). Of course, the actuarial balance is just a shorthand method of describing the gap between projected revenues and benefits, and in practice, the gap could be closed by a combination of revenue increases or benefit cuts.

The baseline scenario for the Social Security actuaries assumes that the total fertility rate will gradually adjust from its 2023 level of 1.62 to a value of 1.9 in 2045, after which it will remain constant. This scenario involves notably higher fertility than the projections of the United Nations and US Census Bureau that were mentioned above, which both have the TFR remaining at or near 1.6. Were the TFR instead to remain constant (specifically, 1.6 from 2045 onward), the 75-year actuarial balance would be −4.49 percent, and were the TFR instead to adjust to a higher level of 2.1 in 2045 and onwards, the actuarial balance would be −3.40 percent (Social Security Administration 2025, Tables V.A1 and VI.D1)

The high and low scenarios for total fertility rate just described—that is, TFR of 2.1 versus 1.6—certainly do not encompass all of the uncertainty regarding how fertility will evolve in coming decades. Many developed countries have TFRs significantly below 1.6, while the United States had fertility well above 2.1 for two decades after World War II. Still this range can usefully serve as a benchmark to illustrate the extent to which low fertility is contributing to the problems of financing support for the dependent elderly. The answer is that fertility is not the dominant factor: between the low- and high-fertility scenarios, the difference in the actuarial balances is 1.09 percent of taxable payrolls, which is only 29 percent of the current actuarial balance. Put differently, reversing low fertility would not on its own fix Social Security, although it could certainly be part of the picture.

One can go further and compare the effect of changing fertility to other adjustments that would affect the actuarial balance similarly. For example, a higher level of immigration is sometimes discussed as a counterbalance to low fertility (see Pritchett's contribution in this symposium). Immigrants tend to arrive young and ready to work—in this sense, they are far more efficient at countering the burden of dependency than newborn babies. Adding an additional 440,000 migrants on top of the 1.25 million annual flow (both legal and illegal) assumed in the baseline of the actuaries would improve the actuarial balance by 0.42 percentage points. Lowering the flow by 420,000 would worsen the balance by 0.46 percentage points (Social Security Administration 2025, Table VI.D3).

Combining these estimates for the effects of fertility and migration, and assuming linearity for small changes around the reported projections, one can readily calculate that the difference between a long-run total fertility rate of 2.1 and 1.6 has roughly the same impact as a change in the annual flow of migrants of 1.07 million people.

This analysis of the fiscal effects of reduced fertility can be extended to include the cost of distortions associated with raising revenue. In the Benefit-Cost Ratio approach of Heckman et al. (2010), assuming that there is no social benefit to the extra Social Security spending beyond the value to the recipient, the cost of $1 of extra Social Security benefits is just $(1 + \phi)$ dollars, where ϕ is marginal deadweight loss, taken to be 0.5. The part of this cost that is not deadweight loss was already incorporated into the calculations above. To give a feel for the deadweight loss: as noted above, the difference in the actuarial balance of Social Security between a future with total fertility rate of 2.1 or 1.6 is 1.09 percent of taxable payrolls. Over the 75-year projection horizon used by Social Security, taxable payrolls average 35 percent of GDP (Social Security Administration 2025, Table VI.G5). Thus, the additional deadweight loss associated with fertility following a lower trajectory would be $1.09\% \times 0.5 \times 0.35 = 0.19\%$ of GDP per year.

This deadweight loss number is relatively small for two reasons. First, as noted above, only part of the actuarial imbalance in Social Security is due to sub-replacement fertility. Second, in the United States, only a relatively small fraction (37 percent) of the gap between consumption and labor income in old age is funded by government transfers (National Transfer Accounts 2016). In many developed countries outside

the United States, the share of consumption of the elderly that is funded by the government is higher (for example, 58 percent in Japan and 80 percent in Italy), projected fertility is lower, and tax rates are higher, meaning that the marginal deadweight loss is higher as well. For all of these reasons, the deadweight loss resulting from sub-replacement fertility would be of considerably greater magnitude.

Other Social Welfare Criteria

I have focused in this essay on the impact of low fertility on consumption per capita. This is, after all, the most frequently used measure of economic success. Nonetheless, in closing it seems worth mentioning other social welfare criteria that might be brought to bear and how their consideration might alter one's view about whether sub-replacement fertility was a problem.

One reason to favor a bigger population is that in such a setting there are more potential friends or mates to choose among, more people who might share one's interest, and so on. Mill (1848) raised this issue, but dismissed it, writing: "The density of population necessary to enable mankind to obtain, in the greatest degree, all the advantages both of co-operation and of social intercourse, has, in all the most populous countries, been attained." A related critique of the per-capita approach starts from the utilitarian imperative to achieve the greatest good for the greatest number. The philosophical approach known as totalist utilitarianism aims to maximize the sum of total welfare, rather than the per-capita level (for discussion in this context, see Spears and Geruso 2025). Most implementations of this approach, taking into account the curvature of the utility function and the productivity of extra people, as well as the utility that people get from being alive rather than just consuming, point to an optimal population size far higher than the current one, and thus suggest that sub-replacement fertility leads to large welfare losses by moving population size in the wrong direction (Adhami et al. 2024).

An alternative philosophical approach, known as long-termism, expands the totalist utilitarian framework by considering the welfare of potential people who could live in the future, without any time discounting. This approach leads to the conclusion that the optimal fertility rate is the one that maximizes the probability of the long-run survival of the human species, given that if the species does survive, the vast majority of human lives will take place far in the future (MacAskill 2022). Many environmentalists argue that in the present setting, low fertility is most conducive to humanity's survival.

Conclusion

It is only over the last half-century that we have started to observe peaceful and prosperous countries with birth rates below the replacement level, and yet today more than half of the world's population lives in countries where this is the case.

The United States, where fertility remained near replacement until the Great Recession, was long an outlier among rich countries.

In this paper, I have attempted to assess the effect of continued sub-replacement fertility on economic wellbeing. My main finding is that this impact, while likely to be negative, is relatively small, even for a large decline in the fertility rate. In addition, this negative impact arrives only after a long adjustment period. For the first four decades, lower fertility provides a demographic dividend of improved economic well-being. Much of the economic cost of population aging currently facing the United States is simply the fading away of the demographic dividend that we reaped from falling fertility in the 1960s and not a result of sub-replacement fertility that emerged after the Great Recession. Conversely, a rise in fertility back to the replacement level, even though it would yield long-term benefits, would reduce the standard of living below what it otherwise would have been for several decades

A society with an older, shrinking population will undeniably feel strange. There will be fewer baby showers relative to funerals, fewer strollers relative to wheelchairs, and houses being torn down to make way for forests rather than the other way around. But society and the economy have gone through plenty of similar dislocations in recent centuries: urbanization, industrialization, the rise of big government, and the entry of women into the labor force, to name just a few. At least as far as my calculations show, a sense of panic about low fertility's economic effects is not justified.

■ *I am grateful to Dario Pietro Matina for research assistance.*

References

Adhami, Mohamad, Mark Bils, Charles I. Jones, and Peter J. Klenow. 2024. "Population and Welfare: Measuring Growth When Life Is Worth Living," NBER Working Paper 31999.

Bloom, David E., David Canning, and Jaypee Sevilla. 2003. *The Demographic Dividend: A New Perspective on the Economic Consequences of Population Change.* Rand Corporation.

Budolfson, Mark, Michael Geruso, Kevin J. Kuruc, Dean Spears, and Sangita Vyas. 2025. "Is Less Really More? Comparing the Climate and Productivity Impacts of a Shrinking Population." NBER Working Paper 33932.

Council of Economic Advisers. 2025. *Economic Report of the President.* US Government Publishing Office.

Cutler, David M., James M. Poterba, Louise M. Sheiner, and Lawrence H. Summers. 1990. "An Aging Society: Opportunity Challenge?" *Brookings Papers on Economic Activity* 21 (1): 1–73.

Eggertsson, Gauti B., Neil R. Mehrotra, and Jacob A. Robbins. 2019. "A Model of Secular Stagnation: Theory and Quantitative Evaluation." *American Economic Journal: Macroeconomics* 11 (1): 1–48.

Ehrlich, Paul R. 1968, *The Population Bomb.* Ballantine Books.

Feenstra, Robert C., Robert Inklaar, and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table." *American Economic Review* 105 (10): 3150–82.

Francis. 2022. "General Audience: Paul VI Audience Hall." Holy See, January 5, https://www.vatican.va/content/francesco/en/audiences/2022/documents/20220105-udienza-generale.html.

Hardin, Garrett. 1995. *Living within Limits: Ecology, Economics, and Population Taboos.* Oxford University Press.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. 2010. "The Rate of Return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94 (1–2): 114–28.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2020. "Land Quality." NBER Working Paper 28070.

Jones, Charles I. 2022. "The End of Economic Growth? Unintended Consequences of a Declining Population." *American Economic Review* 112 (11): 3489–527.

KC, Samir, et al. 2024. "Updating the Shared Socioeconomic Pathways (SSPs) Global Population and Human Capital Projections." International Institute for Applied Systems Analysis Working Paper 24-003.

Kearney, Melissa S., Phillip B. Levine, and Luke Pardue. 2022. "The Puzzle of Falling US Birth Rates since the Great Recession." *Journal of Economic Perspectives* 36 (1): 151–76.

Kotschy, Rainer, and David Bloom. 2023. "Population Aging and Economic Growth: From Demographic Dividend to Demographic Drag?" NBER Working Paper 31585.

MacAskill, William. 2022. *What We Owe the Future.* Basic Books.

Lee, Ronald et al. 2014. "Is Low Fertility Really a Problem? Population Aging, Dependency, and Consumption." *Science* 346 (6206): 229–34.

Malthus, Thomas Robert. 1798. *An Essay on the Principle of Population.* J. Johnson.

Mill, John Stuart. 1848. *Principles of Political Economy.* John W. Parker.

National Center for Health Statistics. 2025. *Division of Vital Statistics. Birth Data 2019.* Public-use data file and documentation. https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/DVS/natality/ (accessed November 3, 2025).

National Transfer Accounts. 2016. "National Transfer Accounts: Data Sheet." https://ntaccounts.org/doc/repository/NTA%20Data%20Sheet%202016.pdf.

National Transfer Accounts. 2025. *Database.* https://www.ntaccounts.org/web/nta/show/Browse%20database (accessed November 3, 2025).

OECD. 2025. *Main Science and Technology Indicators.* https://www.oecd.org/en/data/datasets/main-science-and-technology-indicators.html (accessed November 3, 2025).

Osterman, Michelle J. K., Brady E. Hamilton, Joyce A. Martin, Anne K. Driscoll, and Claudia P. Valenzuela. 2025. "Births: Final Data for 2023." National Vital Statistics Report 74 (1): 1–86.

Social Security Administration. 2022. *Office of the Actuary, Trustees Report.* https://www.ssa.gov/oact/STATS/table4c6_2019_TR2022.html (accessed November 3, 2025).

Social Security Administration. 2025. *The 2025 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds.* SSA.

Spears, Dean, and Michael Geruso. 2025. *After the Spike Population, Progress, and the Case for People.* Simon and Schuster.

United Nations. 2025. *Population Division Data Portal.* https://population.un.org/dataportal/data/indicators/71/locations/840/start/1950/end/2100/table/pivotbylocation?df=f83512ea-f847-4d29-b7f9-3f2dfa58c1e1 (accessed November 3, 2025).

United Nations, Department of Economic and Social Affairs. 2024. *World Population Prospects 2024.* United Nations.

US Census Bureau. 2023. *Methodology, Assumptions, and Inputs for the 2023 National Population Projections.* US Census Bureau.

US Census Bureau. 2025. *Population Division. Annual Estimates of the Resident Population by Single Year of Age and Sex for the United States: April 1, 2010 to July 1, 2019 (NC-EST2019-SYASEXN).* https://www2.census.gov/programs-surveys/popest/tables/2010-2019/national/asrh/nc-est2019-syasexn.xlsx (accessed November 3, 2025).

Weil, David N. 1997. "The Economics of Population Aging." In *Handbook of Population and Family Economics*, Vol. 1B, edited by Mark R. Rosenzweig and Oded Stark, 967–1014. North-Holland.

Weil, David N. 2026. *Data and Code for: "How Much Would Continued Low Fertility Affect the US Standard of Living?"* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E239753V1.

Journal of Economic Perspectives—Volume 40, Number 1—Winter 2026—Pages 47–70

Family Institutions and the Global Fertility Transition

Paula E. Gobbi, Anne Hannusch, and Pauline Rossi

Much of the observed cross-country variation in fertility is consistent with the predictions of classic theories of the fertility transition. That is, countries with higher levels of human capital, higher GDP per capita, or lower child and maternal mortality rates tend to exhibit lower fertility. However, looking at fertility data within countries, larger declines in fertility over the last 60 years are, on average, not associated with greater improvements in human capital, real per capita GDP, or survival rates. Notably, most of sub-Saharan Africa experienced fertility declines smaller than predicted by economic and health progress, while parts of Asia and Latin America experienced declines larger than predicted, and some countries in East Asia even reached record low fertility levels.

To understand why economic and health progress alone fail to account for most of the observed change in fertility over the past half-century, we focus on the role of family institutions, particularly marriage and inheritance customs. We study whether the same institutions can help explain both the stalled fertility transitions in sub-Saharan Africa today as well as the variation in the timing of historical fertility declines across premodern European regions. We also explore

■ *Paula E. Gobbi is Professor of Economics, Université Libre de Bruxelles, Brussels, Belgium. Anne Hannusch is Associate Professor of Economics, University of Bonn, Bonn, Germany. Pauline Rossi is Professor of Economics, Ecole Polytechnique-CREST, Paris, France. All three authors are Research Affiliates, Centre for Economic Policy Research, London, United Kingdom. Their email addresses are paula.eugenia.gobbi@ulb.be, anne.hannusch@uni-bonn.de, and pauline.rossi@ensae.fr.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251460.

whether the diffusion of cultural norms related to religion, educational aspirations, and gender roles can help explain heterogeneous trajectories in the speed and the magnitude of fertility transitions. We then benchmark the quantitative importance of institutional and cultural factors against the effect of economic and health factors. Finally, we investigate how these factors interact to shape fertility transitions. In particular, we analyze whether family institutions, in addition to their direct effect on fertility, also mediate the effect of economic and health factors.

Much of the existing literature examines each factor in isolation, yet no single factor can fully explain all observed fertility transitions or the entire trajectory of a given transition from start to finish. Understanding how these forces interact remains the central challenge. We propose a stylized framework that integrates these factors. As in standard economic models, economic conditions influence fertility decisions. However, these choices operate under constraints determined by the broader environment—health, institutional, and cultural factors—that determine the set of feasible fertility outcomes. We use the theory as a lens to address open questions in the literature: (1) Why does fertility within a country not exhibit a consistent relationship with economic factors over time? (2) Will fertility in sub-Saharan Africa eventually fall below replacement levels once economic forces become sufficiently strong, even without institutional or cultural change? (3) Can policies prevent further fertility decline in East Asia, absent institutional or cultural reform?

What We Know: Economic and Health Factors Matter for Fertility Decline

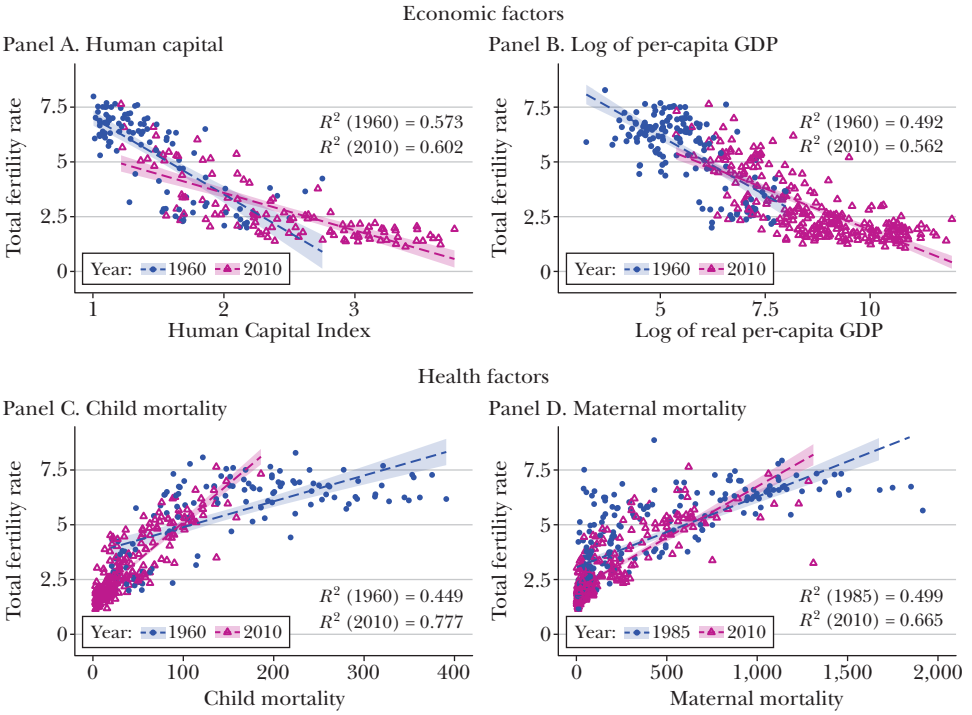
Cross-Country Evidence

The two most prominent explanations of the fertility transition are economic progress (in the economics literature) and health progress (in the demography literature).[1] Figure 1 shows that economic and health factors are indeed strongly associated with total fertility rates across countries at different points in time, where the total fertility rate is the number of children that would be born to a woman if she had children in accordance with age-specific fertility rates observed in a country at a given point in time.

Panels A and B of Figure 1 show that countries with higher levels of human capital and higher GDP per capita tend to exhibit lower fertility rates, both in 1960 and in 2010. In 2010, approximately 60 percent of the cross-country variation in total fertility rates is explained by differences in either human capital or real GDP per capita. Similarly, child and maternal mortality rates are strong predictors of fertility rates, as shown in Figure 1, panels C and D. These health

[1] See Doepke et al. (2023) for a discussion of economic theories of fertility and Guinnane (2011) and Galor (2012) for reviews on demographic transitions.

Figure 1
Evidence from Cross-Country Data in 1960 and 2010



Source: For the *y*-axis: Total Fertility Rate is taken from the World Development Indicator SP.DYN.TFRT.IN (World Bank 2025). For the *x*-axis, panel A is the Human Capital Index from Barro and Lee (2013). Panel B is real GDP per capita from the World Development Indicator NY.GDP.PCAP.CD. Panel C is child mortality as defined as the probability (per 1,000 live births) that a newborn will die before reaching age five, based on the World Development Indicator SH.DYN.MORT. In Panel D, maternal mortality ratio is the estimated number of women who die from pregnancy-related causes while pregnant or within 42 days of pregnancy termination per 100,000 live births, taken from the World Development Indicator SH.STA.MMRT.
Note: Time period: 1960 and 2010. Data on maternal mortality are not available prior to 1985.

variables account for two-thirds to three-quarters of the variation in fertility in the most recent period. High fertility and high mortality tend to coincide in countries with weak reproductive health systems. The effect of economic and health factors on fertility has been documented extensively in the literature; in the following sections, we review important studies that explain the main mechanisms.[2]

[2]For a summary of research studies considering these quantitative effects, see Supplemental Appendix Table A.1.

Economic Factors

The central channel between economic growth and fertility is the quantity-quality trade-off (Becker and Lewis 1973; Galor and Weil 2000). Parents face a trade-off between the number of children they have (quantity) and the resources, such as time and money, they can devote to each child (quality). As the returns to education increase with economic development and the demand for skilled labor rises, parents respond by raising educational investments in each child while having fewer children. To quantify the magnitude of the trade-off, Becker, Cinnirella, and Woessmann (2010) and Bleakley and Lange (2009) exploit plausible exogenous variation in quality in two different contexts: an increase in school enrollment in Prussia in the mid-nineteenth century and an increase in the returns to schooling in the United States in 1910s, respectively. They come to the same conclusion: fertility drops by approximately 20 percent when schooling massively increases.

A second important channel is women's opportunity cost of time (Galor and Weil 1996). As economies develop and incomes rise, the cost of staying at home to raise children and produce domestic goods, rather than participating in the formal labor market, also increases. Using a life-cycle model, Caucutt, Guner, and Knowles (2002) show that US fertility declined by 0.15 children in response to a 12 percent increase women's wage over a decade between 1980 and 1992.

Beyond the well-studied quantity-quality trade-off and opportunity cost of time channels, other economic mechanisms play an important role in explaining why fertility tends to decline with economic development. In traditional economies, children are a source of income and insurance within the family; in modern economies, however, the returns to "quantity" decline for two main reasons.

First, the demand for unskilled child labor falls sharply when economies transition out of agriculture. The need for family labor is a main driver of high fertility in subsistence-farming economies, where child labor is not regulated (Doepke and Zilibotti 2005). For instance, in Burkina Faso, the sustained inflow of remittances from long-standing migration has reduced dependence on subsistence farming and, consequently, the need for child labor, leading to a decline in fertility of about 0.5 children in communities of origin (Dupas et al. 2023). Historically, in the United States in the 1890s, households that switched to manufacturing after an agricultural pest reduced fertility by around 0.25 children compared to those that stayed in agriculture (Ager, Herz, and Brueckner 2020).

Second, the introduction of formal social security systems reduces the need for children as a source of informal old-age insurance. The expansion of social pensions has been shown to substantially reduce fertility, by about 1 to 1.3 children, in contexts like an expansion of social pensions in Namibia in the 1990s (Rossi and Godard 2022) and an equalization of urban and rural pensions in Brazil in 1991 (Danzer and Zyska 2023). Historically, in much of the Western world, the decline in fertility coincided with the introduction of comprehensive social insurance schemes in the late nineteenth century (with the notable exceptions of France and the United States). In a more recent context, Boldrin, De Nardi, and Jones (2015)

develop a macroeconomic model with children as a parental investment in old-age care to study the effect of US and European social security programs on fertility. They find that an increase in program size of 10 percent of GDP is associated with a reduction in fertility of between 0.7 and 1.6 children.

Health Factors

For demographers, the decline in fertility is often seen as a response to the rapid decline in child mortality in the late nineteenth century in Western countries, and after World War II in the rest of the world (Notestein 1952; Preston 1978). There are two channels: the replacement effect captures responses of couples to the loss of a child, whereas the anticipatory or hoarding effect reflects strategies by couples to ensure surviving descendants. Strulik (2004, 2008) develops a theoretical framework in which high child mortality increases fertility, lowers resources per child, and slows human capital accumulation, trapping the economy in a high-fertility, low-education, low-growth equilibrium. Lower child mortality, by contrast, reduces incentives for additional births, raises educational investment per child, and fosters technological progress through human capital accumulation. Despite the prominence of this theory, the causal effect of child mortality remains theoretically and empirically contentious (Wolpin 1997; Doepke 2005; Bar and Leukhina 2010; Baudin 2012). Using a panel of countries from 1900 to 1999, Herzer, Strulik, and Vollmer (2012) find that a 1 percent decline in child mortality reduces fertility by only 0.14 percent in the long run. This modest elasticity implies that the decline in child mortality alone cannot account for the magnitude of the fertility transition.

Maternal mortality is another often-mentioned health factor. In theory, improvements in maternal health should raise fertility by reducing the physical cost of childbearing, but they may in fact lower fertility in the long run if they increase the returns to female human capital accumulation. The sharp reduction of maternal mortality in the United States between 1930 and 1950 led to a short-run increase in fertility, followed by a long-run decrease twice as large as the short-run response (Albanesi and Olivetti 2014).

More generally, child and maternal mortality rates can be seen as proxies for the quality of reproductive health systems and the availability of modern contraceptive methods. The effect of modern birth control on fertility is debated, given that traditional methods of contraception have long been available (for instance, Cinnirella, Klemp, and Weisdorf 2017), and that modern contraception historically followed, rather than preceded, fertility decline in the Western world (Pritchett 1994; Hartmann 1997). The contribution of family planning to contemporary fertility transitions in low- and middle-income countries varies across contexts. A review of the empirical evidence using the introduction of family planning programs as natural experiments reports reductions in fertility between 5 percent and 35 percent depending on the program (Miller and Babiarz 2016). The most impactful programs combine access to contraception, improvements in child health, and intensive communication campaigns.

The Puzzle: Economic and Health Progress Are Not Enough to Account for All Fertility Transitions

Within-Country Evidence

The correlation between economic and health factors weakens when analyzing changes within countries over time. Figure 2 uses the same data as Figure 1, but now focuses on within-country changes. Colors and shapes denote each country's region. Panels A and B of Figure 2 show that when looking at within-country data, larger declines in fertility between 1960 and 2010 are, on average, only weakly associated with greater improvements in human capital or real per capita GDP. Similarly, panels C and D of Figure 2 show no strong association between changes in child or maternal mortality and changes in fertility within countries over time. In fact, changes in economics and health variables alone explain at best 7 percent of the within-country variation in fertility over this period.

Many countries in East Asia and Pacific (□), South Asia (✳), Latin America and the Caribbean (■), and Middle East and North Africa (○) experienced substantial fertility declines without corresponding increases in income, human capital, or health.

Conversely, most countries in sub-Saharan Africa (●) experienced significant health progress and to some extent economic progress without a substantial decline in fertility. European (▲) and North American countries (△) experienced strong economic progress and only limited fertility declines. Taken together, these results suggest that countries followed distinct fertility transition paths over the past 60 years.

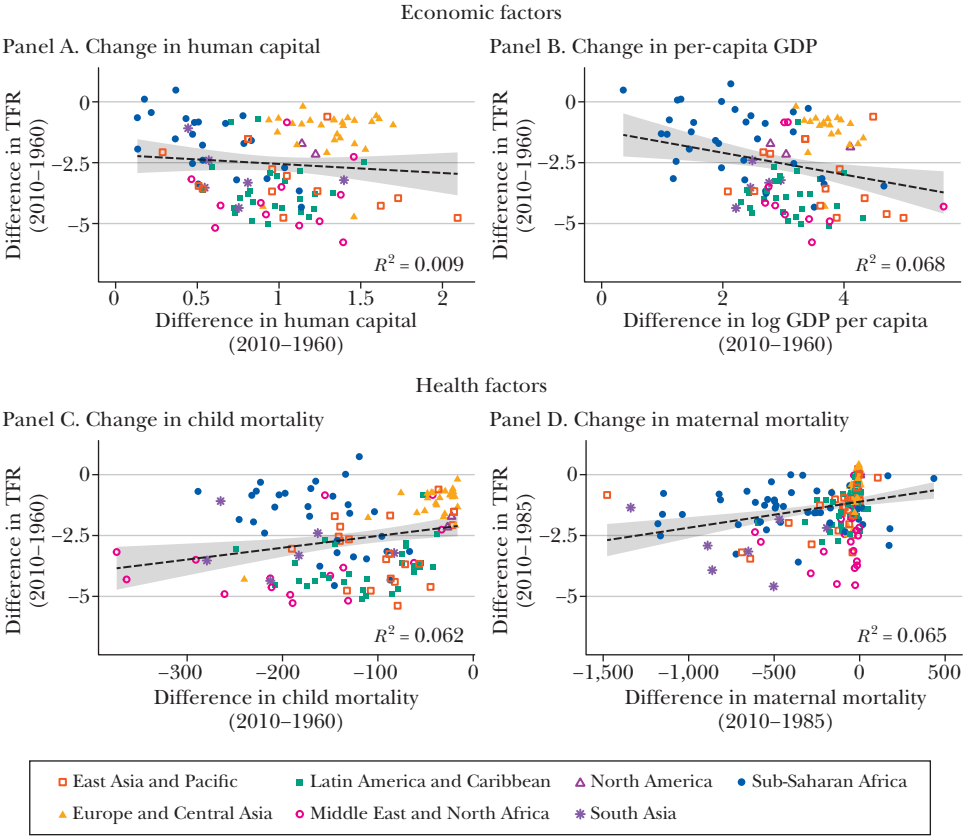
Heterogeneous Transition Paths

More generally, when examining fertility transition paths country by country, several puzzles emerge. The most debated case is that of France, the first country where fertility began to decline. The decline started in the mid-eighteenth century, one century before the rest of Europe and the United Kingdom, and well before the onset of modern economic growth and the decline in mortality. The French trajectory is therefore completely at odds with classic theories of the fertility transition. A similar puzzle arises in the United States, another well-known forerunner, where the initial trigger of the decline remains a subject of debate.

Other exceptional cases are East Asian countries, notably Japan, South Korea, and China, where fertility declined so rapidly in the second half of the twentieth century that they reached ultra-low levels of fertility. Fertility levels in East Asia not only converged toward Western levels but continued to fall even in countries, such as China, that had not yet caught up in terms of economic development. None of the classic theories of fertility transition can predict how low fertility will drop.

Finally, another frequently cited exception is sub-Saharan Africa, where the decline has been particularly late and slow, and future trajectories remain difficult to predict. For example, comparing the region's two largest countries, Nigeria and Ethiopia, the decline has been slower in Nigeria, despite its higher income level. This raises the question of whether sub-Saharan Africa is truly exceptional, and if so,

Figure 2
Evidence from Within-Country Changes over Time



Source: For the *y*-axis: Total Fertility Rate is taken from the World Development Indicator SP.DYN.TFRT.IN (World Bank 2025). For the *x*-axis, panel A: Human Capital Index from Barro and Lee (2013). Panel B: real GDP per capita from the World Development Indicator NY.GDP.PCAP.CD. Panel C: Child mortality is defined as the probability (per 1,000 live births) that a newborn will die before reaching age five, based on the World Development Indicator SH.DYN.MORT. Panel D: Maternal mortality ratio is the estimated number of women who die from pregnancy-related causes while pregnant or within 42 days of pregnancy termination per 100,000 live births, taken from the World Development Indicator SH.STA.MMRT.
Note: Time period: 1960 and 2010. Data on maternal mortality are not available prior to 1985.

what makes it so. To address this question, we turn to the role of family institutions and cultural norms.

Missing Pieces: Family Institutions and Culture

How can we explain the circumstances under which economic and health factors alone fail to account for much of the observed change in fertility over the

past half-century? We highlight one additional channel: the role of *family institutions*. We focus on two types of long-standing family institutions: marital structure, referring to the prevalence of monogamous versus polygamous marriage, and inheritance customs, namely partible versus impartible inheritance. Under impartible inheritance, property, assets, or wealth are passed on to a single heir, typically the eldest son. This system was common in historical England and in several regions of continental Europe before the adoption of harmonized national civil codes, and it still exists in parts of sub-Saharan Africa where family law falls under customary law. By contrast, partible inheritance divides property equally among multiple heirs.

Family Institutions: Marriage and Inheritance

The role of family institutions in shaping aggregate fertility has long been discussed by historians, demographers, and sociologists (Todd 1984; Lesthaeghe 1989). In particular, customary law related to marriage and inheritance may help explain both pretransition differences in fertility within countries and variation in the timing of the transition's onset. Recent empirical studies have begun to shed light on the magnitude of these effects.

In Western Europe, both marriage and inheritance are shown to be quantitatively important in explaining why fertility was already low during Malthusian times, and why fertility began to decline before the transition to modern growth. The "European Marriage Pattern," characterized by late marriage and high rates of celibacy, implied for centuries that women typically married after age 23–24, and 10 to 15 percent never married at all. In contrast, female marriage was early and universal in most other parts of the world (Hajnal 2017). Because later marriage reduced the number of women exposed to pregnancy risk, the European Marriage Pattern is estimated to have reduced birthrates by 20 to 40 percent between the fourteenth and nineteenth centuries (Voigtländer and Voth 2013; Perrin 2022).

Customary inheritance laws, and in particular the equal division of land between children, created incentives to restrict fertility to avoid land fragmentation across generations. Regional variation in such laws is correlated with fertility differences between French regions during the early eighteenth century. The general adoption of equal partition after the French Revolution contributed to France's early fertility decline, reducing completed fertility by around 0.5 children (Gay, Gobbi, and Goñi forthcoming).[3]

In the same vein, marriage and inheritance institutions can explain the high levels of fertility and the stalled demographic transition in modern sub-Saharan Africa. The region exhibits persistently high rates of polygamy, which sustains a high level of fertility through several mechanisms. Polygamy is associated with (1) early and universal female marriage, which maximizes women's exposure to pregnancy; (2) high bride-price, which incentivizes parents to have many daughters (Tertilt

[3]Similarly, in the seventeenth and eighteenth centuries, the British nobility often used marriage settlements to entail land and thus prevent the (solo) heir from breaking up the family estate. These inheritance practices increased fertility through a reduction in childlessness (Gobbi and Goñi 2021).

2005); and (3) rivalry between co-wives, which motivates each wife to have more children than the others (Rossi 2019). A quantitative model predicts that banning polygamy would reduce fertility by 40 percent (Tertilt 2005). Cousin marriage (also called within-kin or endogamous marriage), also a prevalent institution in parts of Africa and the Middle East, may also raise fertility by reducing search frictions in the marriage market and limiting wealth fragmentation from exogamous (that is, outside-group) unions (Lesthaeghe 1989).

The customary inheritance laws across countries of Africa tend to favor impartible inheritance, enabling the transfer of family land to a single heir and removing the incentives to limit the number of heirs. Today, belonging to an ethnic group with a tradition of impartible inheritance increases fertility by around one child compared to neighboring ethnic groups (Fontenay, Gobbi, and Goñi 2025). Moreover, customary laws excluding widows from inheritance rights increase women's reliance on their children for economic security. In these contexts, high fertility reflects women's strategies to mitigate risks related to divorce or widowhood. For example, Lambert and Rossi (2016) show that when women in Senegal are married to a man with children from previous wives, which reduces the probability of a substantial inheritance, they have more sons. As a consequence, granting widows a fair share of the husband's bequest should weaken these strategies. In Namibia, a 2008 reform improving widows' inheritance rights reduced the annual birthrate by 24 percent, equivalent to a reduction in completed fertility by one child (Sage 2025).

Recognizing that some institutional contexts can generate incentives for women to desire larger families is important for understanding recent empirical patterns that may otherwise seem puzzling. For example, experimental interventions aimed at promoting female economic and reproductive empowerment involving business training and land titling caused an *increase* in fertility in Togo, Ethiopia, Benin, and Ghana (Donald et al. 2024), as well as in Tanzania (Berge et al. 2022). Similarly, experimental evidence involving husbands in family planning interventions has been shown to reduce contraceptive take-up among monogamous households, but to *raise* contraceptive take-up in polygamous households in Burkina Faso (D'Exelle et al. 2023). These findings are consistent with the idea that women want many children, which makes sense when family law is designed to reward high fertility.

Other institutions play an important role at later stages of the fertility transition, in particular by shaping the career-family trade-off faced by mothers. This trade-off is absent when women's economic opportunities are limited by legal restrictions. As women's legal and economic rights expand, however, the trade-off becomes salient. When childcare institutions remain underdeveloped, fertility typically declines. For instance, comparisons of neighboring counties that have a state-level border running between them shows that the improvement in women's legal and economic rights in some states and not others during the late nineteenth century in the United States reduced fertility by around 7 percent or 0.2 children (Hazan, Weiss, and Zoabi 2023). In contemporary settings, the rapid improvement

in female career prospects alongside a high burden of domestic work has been proposed as an explanation for East Asia's exceptionally low fertility levels (Goldin 2025). The emergence of childcare institutions, whether market- or state-based, eases the trade-off and women can combine having a family and a career, which implies both higher levels of fertility and higher levels of female labor force participation. This was shown by d'Albis, Gobbi, and Greulich (2017) using cross-sectional data for 2011 across European countries with different levels of childcare, by Bar et al. (2018) looking at the effect of marketization of childcare on high- and low-income US women from 1980 to 2010, and by Hazan, Weiss, and Zoabi (2021) using changes in relative childcare costs for US women with different levels of education from 1980 to 2020. In addition, labor market institutions can influence the career-family trade-off. Guner, Kaya, and Sánchez-Marcos (2024) develop a life cycle model to show that temporary contracts or split-shift jobs reduce fertility in Spain. Removing these features, in combination with childcare subsidies, could raise fertility in Spain by 0.22 children.

Cultural Factors

Cultural factors, or norms, will also affect how fast and how low fertility drops, operating through two main mechanisms: (1) directly, by shaping norms about the ideal family size; and (2) indirectly, by influencing the acceptability of birth control and perceived costs of child rearing.

A central hypothesis in demography is that limiting births was a cultural innovation, first observed in France, that spread through social interactions across Europe and European offshoots (Coale and Watkins 1986; Bongaarts and Watkins 1996). Fertility declines often diffused to culturally or geographically close communities (Delventhal, Fernández-Villaverde, and Guner 2024). In France, the decline radiated from low-fertility regions, in particular Paris, to the rest of the country via internal migration (Daudin, Franck, and Rapoport 2019). Across Europe, the decline propagated from French-speaking regions to culturally similar communities before reaching more distant ones, again with migration playing a central role (Spolaore and Wacziarg 2022; Melki et al. 2024). In the English-speaking world, the sharp fertility decline in Britain in the late nineteenth century was mirrored among British migrants in Canada, the United States, and South Africa (Beach and Hanlon 2023). Among second-generation American women, higher ancestral-country fertility predicts about 0.4 more children (Fernández and Fogli 2009). Similarly, in China, fertility reductions imposed on the majority ethnic group, the Han Chinese, in the 1970s spilled over to culturally close minority groups, despite their exemption from birth quotas (Rossi and Xiao 2024).

At the start of the fertility transition, cultural attitudes toward the "morality" of birth control were pivotal. In France, where secularization was already advancing in the eighteenth century, weakening religious influence helps explain why the transition began there first. Within France, regions with high secularization experienced fertility transitions up to a century earlier and completed fertility about one child lower than fully religious areas (Murphy 2015; Blanc 2023), with similar patterns

in Belgium (Lesthaeghe 1977). In addition, Perrin (2022) argues higher gender equality and women's agency, in combination with secularization, contributed to the decline. In Britain, the break in the fertility trend around 1877 coincided with a high-profile trial, in which Charles Bradlaugh and Annie Besant published a book making a case for the right to choose family size and offering some basic information about contraception. Bradlaugh and Besant knew they were very likely to be arrested for doing so, and then used their trial to publicize the benefits of birth control (Beach and Hanlon 2023). A modern parallel comes from Brazil, where the broadcast of soap operas (*novelas*) influenced fertility choices by promoting smaller family size norms (La Ferrara, Chong, and Duryea 2012). In the United States, Kearney and Levine (2015) show that a reality show on teenage childbearing led to a 4.3 percent reduction in teen births.[4]

This bottom-up spread of birth control norms sharply contrasts with the post–World War II transition in other parts of the world, where population control policies, ranging from mildly paternalistic to strongly coercive, were often imposed from the top down. Such policies are aimed at establishing a radically different family size norm over a short period of time (De Silva and Tenreyro 2017) and explain why post–World War II fertility declines were much faster than the gradual historical transitions. One exception is sub-Saharan Africa, where such policies often clashed with deep-rooted religious and traditional customs that emphasize the role of ancestral lineage and where extended family members often influence a couple's fertility decisions (Caldwell and Caldwell 1987). Today, ethnic groups placing high value on the perpetuation of family lineage have fertility rates higher by 0.5 to 1 child compared to others (Álvarez-Aragón 2025).

Toward the end of the fertility transition, cultural factors help explain the stark differences across modern economies. While many countries in East Asia have reached a "lowest-low" fertility level, several Western countries remain near replacement levels. In contexts with high labor market inequality or strong social norms around educational attainment, an "education fever" emerges, as parents compete in terms of resources spent per child (Mahler, Tertilt, and Yum forthcoming). This competition raises the cost of children and hence how many children can be afforded. Using a quantitative model calibrated to South Korea, Kim, Tertilt, and Yum (2024) show that fertility would be 28 percent higher absent status externalities in education. Rising education also affects the marriage markets. In much of East Asia, women marry later or remain unmarried if suitable partners are scarce, and given that out-of-wedlock births are rare, this trend amplifies fertility decline. In China, the growth in the educated population is estimated to explain half of the drop in marriage rates, partly because educated women rarely marry less-educated men (Rossi and Xiao 2026). Finally, Kearney and Levine (2025) attribute the fertility decline in high-income countries to rising childlessness among cohorts whose priorities shifted away from parenthood.

[4]However, see Jaeger, Joyce, and Kaestner (2020) for a critique of Kearney and Levine (2015)'s findings.

New Lessons and Puzzles

There is considerable micro- and macroeconomic evidence on the quantitative importance of economic, health, institutional, and cultural factors in shaping fertility (for a summary of this literature, see Supplemental Appendix Table A1). Our reading of the literature is that the effects of family institutions and culture on fertility are comparable, and in some cases even larger, than the effect of economic and health factors that have been the main focus of the economics and demography literatures.

But how do all of these factors interact? Existing evidence suggests that the effects of economic and health factors depend on the social context. For instance, the magnitude of the quantity-quality trade-off is estimated to be much smaller in sub-Saharan Africa than it was historically in Western countries. Vogl (2025) shows that fertility decline is only weakly associated with the educational progress of children within countries or regions, while Collins, Guarnieri, and Rainer (2025) estimate that free primary education in recent decades reduced fertility by only 0.1 child, or 4 percent. The impact of family planning interventions appears similarly limited. Randomized controlled trials in three sub-Saharan African countries find that free access to contraception has a negligible effect on births (Desai and Tarozzi 2011; Ashraf, Field, and Leight 2013; Dupas et al. 2025).

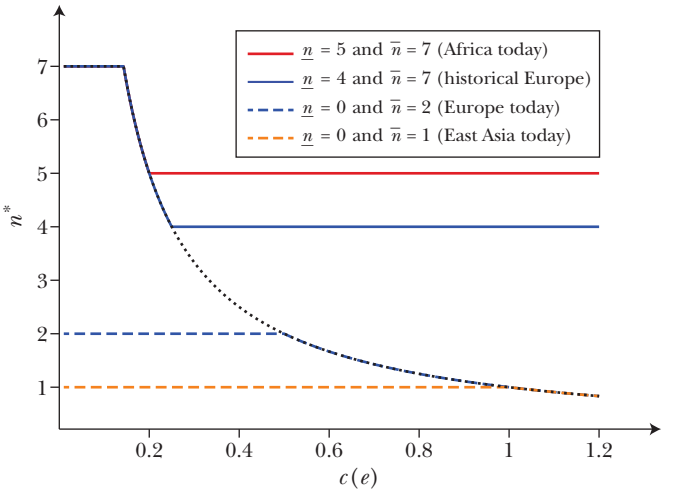
Similarly, health factors interact with economic conditions. For example, Cervellati and Sunde (2011) exploit cross-country variation in mortality reductions due to epidemiological changes and find that a 1 percent increase in life expectancy is associated with a 1.4 percent reduction in fertility, but only after the onset of the demographic transition, not at earlier stages of economic development. Such findings suggest that we are still far from a unified theory of the fertility transition that fully captures the interactions between all factors. In the next section, we propose a first step in that direction.

An Extended Model of Fertility Transitions

We develop a framework in which fertility choices respond to economic incentives. The influence of these economic incentives, however, can be mediated by prevailing family institutions, cultural norms, and health conditions. Institutions include marriage and inheritance. Culture relates directly to fertility through the ideal family size, and indirectly through religious beliefs, attitudes towards sex and contraception, educational expectations (for example, competitive schooling environments), or gender roles in society. Finally, health-related factors encompass child and maternal mortality, access to contraception, and the availability of infertility treatments. We illustrate the model with examples of settings where economic factors alone have had limited impact on fertility.[5]

[5] For an algebraic presentation of the model, please refer to Supplemental Appendix A.1.

Figure 3
Fertility Transitions Paths Given Different Fertility Bounds $\{\underline{n}, \bar{n}\}$



Source: Author's own calculations.
Note: This figure plots the path of optimal fertility n^* as a function of $c(e)$, for a given set of lower and upper fertility bounds, $\underline{n}$ and $\bar{n}$. These bounds are chosen to reflect prevailing fertility rates in different historical and modern contexts. Within these bounds, fertility is determined by $1/c(e)$. See Supplemental Appendix Equation A.1 for the solution to the simplified model.

Fertility Transition Paths

Figure 3 illustrates different fertility transition paths captured by the model. The vertical axis shows the equilibrium fertility level in a given economy. The horizontal axis reflects the broad economic costs of having a child, which include the opportunity cost of child-rearing relative to being in the labor market, returns to human capital, the ability to accumulate assets, and the economic value of child labor (when children are young) and support from children (when the children have become adults) for parents in their old age. One can think of the horizontal axis as net costs, also taking into account the economic benefits of having children.

The transition paths highlight that fertility decisions are made within a broader societal context, shaped by factors that are often not fully incorporated into standard economic models. Specifically, a combination of family institutions, culture, and health technology impose bounds on fertility choices. Between these bounds, fertility responds to economic factors in the way predicted by standard economic theories of fertility. However, when fertility is constrained by prevailing family institutions, culture, or health factors, changes in economic incentives no longer influence fertility behavior.

This framework encompasses several classes of models that analyze fertility transitions. Demographers often focus on how upper and lower fertility bounds change over time when health factors, such as child mortality or the availability of

contraception, vary. Social scientists typically stress the role of family institutions in shaping these fertility bounds. For example, institutions encouraging late female marriage tend to relax the lower bound and tighten the upper bound, because women are married during a shorter span of their reproductive years. Finally, diffusion models of the fertility transition focus on how changes in the bounds propagate through changes in cultural factors.

Interpretation and Examples

To illustrate the intuition behind the model, we compare transitions in historical Europe and sub-Saharan Africa, as well as modern Europe and East Asia. We begin with a comparison of historical Europe and contemporary sub-Saharan Africa, which both represent early stages of fertility decline. We assume that both settings share the same upper bound on fertility, interpreted as the biological maximum number of children that a substantial population of women can have, set at $\bar{n} = 7$. The two settings differ, however, in their health environment, institutions, and norms that lead to different lower bounds: the minimum achievable number of children. For illustration, the lower bound is set to $\underline{n} = 4$ in historical Europe (blue solid line) and to $\underline{n} = 5$ in sub-Saharan Africa (red solid line). The higher $\underline{n}$ in sub-Saharan Africa may stem from family institutions such as polygamy or impartible inheritance, strong pronatalist norms, or higher mortality rates.

This comparison delivers several important insights. First, institutions can determine how low fertility falls, even in Malthusian, preindustrialized contexts. For example, the European Marriage Pattern and equal inheritance in early Europe facilitated lower fertility compared to regions at the same stage of economic development where these institutions were absent. Conversely, polygamy and impartible inheritance sustain higher fertility in some sub-Saharan African regions today, despite facing similar economic conditions.

Second, in preindustrial times, when the costs of raising children are relatively low, fertility oscillated between $\underline{n}$ and $\bar{n}$, as shown in Figure 3, in response to short-run fluctuations in economic conditions. If the environment shifts, due to changes in institutions, health conditions, or social norms, a lower $\underline{n}$ can be achieved, for instance, through the adoption of partible inheritance or monogamous marriage. Such a shift gives greater scope to economic factors in shaping fertility decisions.

Third, the timing of changes in the four factors affects fertility transition paths. Economic development leads to a sustained fertility decline only once institutional, health, or cultural constraints are not binding. A similar argument is put forward in Spolaore and Wacziarg (2022), who argue that economic forces reduce fertility only when the social stigma has declined enough to no longer anchor fertility. In France, although the transition to lower fertility spanned over two centuries (from the mid-eighteenth to early twentieth century), fertility was already low by the mid-nineteenth century, following deep cultural change through secularization and institutional shifts, such as the harmonization of legal institutions. In other words, the initial decline in fertility reflected a downward shift in $\underline{n}$. Once this constraint eased, economic factors related to industrialization drove the second stage of the

transition, allowing n^* to move along the black dotted line as $c(e)$ increased. By contrast, in sub-Saharan Africa, rising $c(e)$ from investments in education and urbanization have yet to lower fertility. Here, cultural norms and institutional structures, such as legal pluralism and customary family law, form a tightly intertwined set of slow-changing constraints. Differences in the sequence of how the four factors change thus help explain timing differences in the onset of fertility decline.

We now turn to the interaction of these factors in the later stages of fertility decline by comparing fertility patterns in Europe and East Asia today. At the end of the fertility transition, the lower bound is no longer binding, $\underline{n} = 0$, because women can socially choose to remain childless. However, these regions differ in their upper bounds, interpreted as the maximum number of children achievable given, for example, the availability of suitable partners or access to infertility treatments. We assume $\bar{n} = 2$ in Europe (blue dashed line) and $\bar{n} = 1$ in East Asia (yellow dashed line). When the cost of children is moderately high (say, $c(e) < 0.5$ in Figure 3), each economy is capped at a different upper bound: 2 in Europe and 1 in East Asia. As the cost of children rises further, both economies converge toward zero fertility.

During the European fertility transition, social change was gradual and spontaneous, with norms around ideal family size evolving slowly, allowing labor and marriage markets time to adapt. As a result, women now combine careers with motherhood without excessive postponement, and when delays occur, they have access to subsidized assisted reproductive technology. Two children thus remains the ideal family size reported by most individuals in many, though not all, European countries.[6] These patterns can be interpreted as a downward shift in $\bar{n}$, followed by movement along the black dotted line, where $\bar{n}$ no longer binds. By contrast, a number of East Asian countries experienced strict population policies that imposed new family size norms, such as China's one-child family. This rapid, top-down social change coincided with worsening sex ratios, education levels that rose more quickly for women than for men, and persistent female hypergamy (women marrying men with higher socioeconomic status). These shifts had profound and lasting consequences for the marriage market. In this context, $\bar{n}$ in Figure 3 declined more rapidly than in the European case, constraining today's fertility at lower levels.

This conceptual framework offers a way of approaching several open questions in the fertility literature. Why is there no consistent relationship between declining fertility and rising human capital or per-capita GDP within countries over time? We argue that fertility responses can be constrained by institutional, health, and cultural factors that mitigate the effect of economic factors on fertility. Will fertility in sub-Saharan Africa eventually converge and drop below replacement levels once economic forces become strong enough? In the absence of institutional reforms related to land property, marriage, or inheritance customs, this appears unlikely. Can policy prevent fertility in East Asia from dropping below

[6]See Mahler, Tertilt, and Yum (forthcoming) for a recent discussion on the determinants of ultra-low fertility in OECD countries.

one child per woman? In the absence of institutional or cultural changes that raise $\overline{n}$, policies that only target the cost of having children are unlikely to succeed.

Solving the Puzzle: Economic Development Matters, but under Specific Family Institutions

The model delivers one testable implication: economic development should matter more for fertility decline when family institutions are already favorable to small families. To identify prevailing family institutions, we use data from the Ancestral Characteristics database (Giuliano and Nunn 2018). These data provide worldwide country-level measures for the share of the current population that has a given preindustrial ancestral characteristic.[7] The Ancestral Characteristics database is based upon the Ethnographic Atlas, which is an ethnicity-level database covering over 1,200 preindustrial societies (Murdock 1967). This allows us to classify countries as either (historically) primarily monogamous or polygamous, and as adhering mainly to partible or impartible inheritance systems.[8]

To assess the explanatory power of economic and health factors conditional on family institutions, we focus on two time periods: 1975–1985 and 2013–2023. We average all annual country-level variables within each decade to account for data gaps and then compute the differences between periods. We regress changes in total fertility rate (TFR) on changes in economic variables (log of real per-capita GDP and secondary school enrollment) and health variables (maternal and child mortality). Using the estimated coefficients, we predict fertility changes and compare them to actual changes.

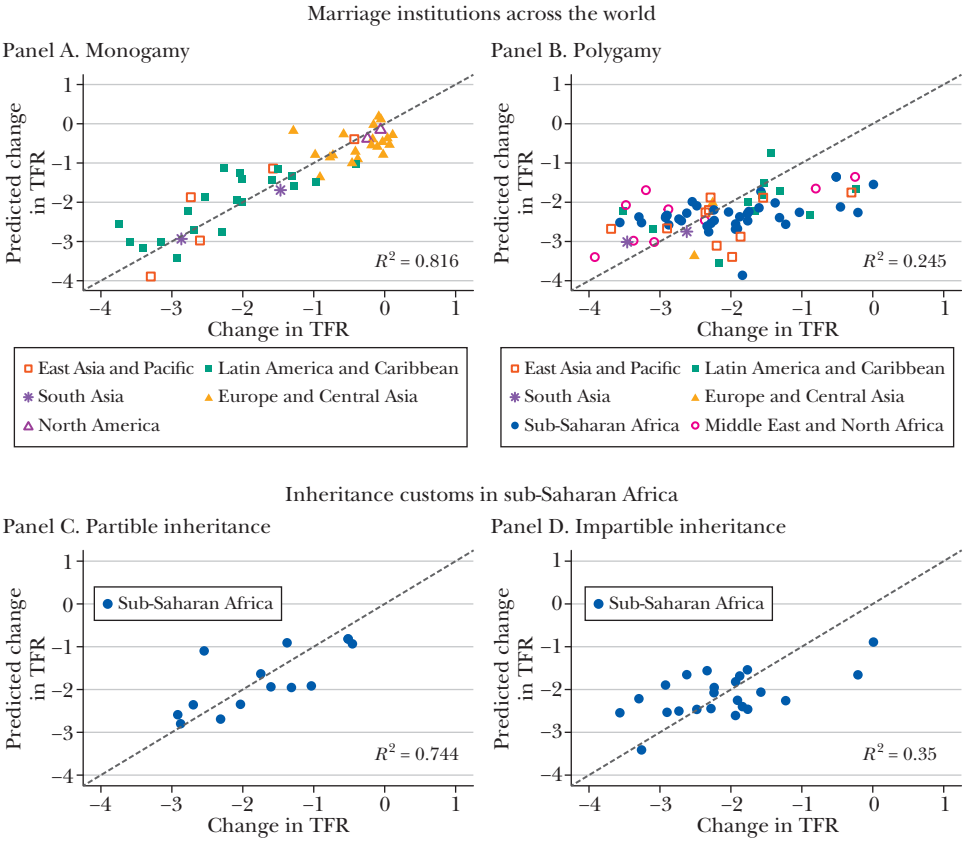
Panels A and B of Figure 4 show the results when we distinguish countries by marital institutions. Strikingly, economic and health factors explain a large share of the fertility changes over time within countries where monogamy is prevalent. In these settings, over 80 percent of the variation in fertility decline is accounted for by economic and health changes. This stands in stark contrast to our earlier findings in Figure 2, where no such association emerged when family institutions were not considered. Thus, when focusing exclusively on monogamous countries, the data appear largely consistent with classic theories of fertility decline. In contrast, in polygamous countries, the explanatory power of economic and health factors drops sharply, explaining less than 25 percent of the observed fertility changes.[9]

[7] Giuliano and Nunn (2018) link ancestral characteristics to contemporary countries by assigning ethnographic traits to language groups and combining their geographic distribution with high-resolution population data. The resulting spatial distribution of ancestral traits is aggregated using modern country borders to construct country-level measures of average ancestral characteristics.

[8] Specifically, we use the variables v9 and v75 for the prevalence of monogamous versus polygamous marriages and for partible versus impartible inheritance rules, respectively. We categorize a country as having monogamous or partible ancestral institutions if more than 50 percent of its population descends from groups that historically practiced monogamous marriage or partible inheritance.

[9] The detailed presentation of the regression findings in Supplemental Appendix Figure A.1 further shows that most of the explanatory power comes from economic variables, with health variables contributing little in polygamous contexts.

Figure 4
Family Institutions: Marriage Institutions and Inheritance Customs



Source: Giuliano and Nunn (2018) and World Bank (2025).
Note: Time period: 1975–1985 and 2013–2023. We estimate a linear regression of changes in total fertility rate on changes in economic factors (log GDP per capita, secondary school enrollment) and health factors (maternal mortality, child mortality). All variables are averaged over the two decades (1975–1985 and 2013–2023), and changes are computed as differences between these averages. Using the estimated coefficients, we predict changes in total fertility rate and plot them against the observed changes.

Panels C and D of Figure 4 focus on the second type of family institution: inheritance customs. Here, we focus our analysis on countries of sub-Saharan Africa, where the Ethnographic Atlas offers particularly rich data capturing ethnic variation in inheritance customs within countries. Outside the countries of Africa, most other countries harmonized their inheritance laws during the nineteenth and early twentieth centuries, meaning that ancestral practices no longer reflect the legal institutions in place during the second half of the twentieth century. Again, we observe that the explanatory power of economic and health factors depends on the underlying family institution. In countries where partible inheritance is the norm, these factors explain nearly 75 percent of fertility changes, a magnitude similar to

monogamous countries. However, in countries with impartible inheritance, the explanatory power declines to around 35 percent.[10]

Altogether, Figure 4 highlights that the pace of fertility decline cannot be understood without reference to family institutions. Economic and health improvements predict fertility decline only in contexts where noneconomic factors already favor smaller families, such as monogamous marriage systems or partible inheritance. These results underscore that economic modernization alone is insufficient to trigger rapid fertility decline: its effects critically depend on preexisting social and institutional conditions that shape how families respond to economic incentives.

The evidence presented thus far is not intended to establish causality. Family institutions are shaped by deep-rooted determinants, such as climatic and geographic conditions as well as the characteristics of original tribes and early settlers,[11] and these determinants may also mediate the effect of economic development. The purpose of our analysis is to highlight the strong interplay between family institutions, health, and economic factors in shaping fertility outcomes. In the long-run, these interactions are even stronger than illustrated here, because institutions and culture are themselves endogenous to economic and health factors. For instance, the transition from polygamy to monogamy can be explained by changes in the distribution of income. With economic development, human capital accumulation, social mobility, and redistribution, monogamy becomes more attractive for both women and men, because the pool of marriageable men increases and the returns to large families decline (de la Croix and Mariani 2015). Another example involves the emergence of the European Marriage Pattern, which can be traced to a single dramatic event: the Black Death of the mid-fourteenth century. The resulting increase in the land-to-labor ratio is thought to have shifted agricultural production from grain to livestock, an activity in which women had a comparative advantage. This improved female employment prospects and contributed to later marriages (Voigtländer and Voth 2013). Institutional change may therefore result from the combination of health shocks and economic transformations.

Conclusion and Policy Implications

Determining the optimal population size and growth rate is far from straightforward (Golosov, Jones, and Tertilt 2007; de la Croix and Doepke 2021), and some philosophers even argue it is impossible (Parfit 1984; Arrhenius 2000). Historically, fears of "population explosion" or "population decline" have motivated coercive antinatalist or pronatalist policies in many countries, though such approaches have

[10] Supplemental Appendix Figure A.2 further distinguishes the effect of economic and health factors.

[11] See Fernández and Fogli (2006) and Haddad (2023) on cultural persistence and Jones (2003, pp. 15–21) on geographic conditions: comparing the European Marriage Pattern with Asian societies, Jones argues that the recurrence of natural disasters in the Indian subcontinent encouraged early marriage as a strategy to achieve high birthrates, providing demographic insurance against recurring environmental shocks.

increasingly come under criticism (Hartmann 1997; De Silva and Tenreyro 2017). Today, most state interventions instead focus on narrowing the gap between desired and actual fertility. The United Nations, for example, has recommended investing in reproductive health and family planning to "enable women and couples to achieve their desired family size" (United Nations 2020). In practice, governments typically aim to keep fertility close to the replacement rate, around two children per woman.

Given the complex interplay among economic, health, institutional, and cultural factors highlighted in this essay, interventions targeting any single dimension seem unlikely to generate large changes in fertility. Mitigation policies aimed at directly influencing fertility—whether to reduce it in high-fertility countries or raise it in low-fertility ones—tend to have only modest effects. In high-fertility settings, current strategies focus on improving access to contraception; as discussed earlier, such measures have limited impact when desired fertility remains high, though they are essential to facilitate the decline once couples are willing to control births. In low-fertility contexts, policies such as financial incentives, childcare provision, and infertility treatments have similarly modest effects (see the review by Bergsvik, Fauske, and Hart 2021), but can help slow the pace of decline.

Addressing multiple contributing factors simultaneously poses a major challenge, and there is little knowledge on when it is optimal to push one factor versus another, making a policy-induced global-fertility rebound improbable in the near future. In light of this, adaptation policies—those aimed at managing the wide array of consequences of ongoing fertility decline—seem unavoidable.

■ *This research was cofunded by the German Research Foundation (CRC-TR-224 project A3) and by the European Union (ERC projects IDED and P3OPLE). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. We are grateful to the Managing Editor, Timothy Taylor, the Editor, Heidi Williams, and Coeditors Jeffrey Kling and Jonathan Parker for detailed feedback on our first draft. We also thank participants to the conference on fertility and demographic change at UCLA, as well as Philip Ager, David de la Croix, Claudia Goldin, Joanne Haddad, Juliana Jaramillo Echeverri, Michèle Tertilt, Tom Vogl, Romain Wacziarg, Minchul Yum, and Anson Zhou for helpful comments.*

References

Ager, Philipp, Benedikt Herz, and Markus Brueckner. 2020. "Structural Change and the Fertility Transition." *Review of Economics and Statistics* 102 (4): 806–22.

Akerlof, George A. 1997. "Social Distance and Social Decisions." *Econometrica* 65 (5): 1005–27.

Albanesi, Stefania, and Claudia Olivetti. 2014. "Maternal Health and the Baby Boom." *Quantitative Economics* 5 (2): 225–69.

Álvarez-Aragón, Pablo. 2025. "Paths *to* the Rainforests: Ancestral Beliefs and Fertility in Sub-Saharan Africa." Unpublished.

Arrhenius, Gustaf. 2000. "An Impossibility Theorem for Welfarist Axiologies." *Economics and Philosophy* 16 (2): 247–66.

Ashraf, Nava, Erica Field, and Jessica Leight. 2013. "Contraceptive Access and Fertility: The Impact of Supply-Side Interventions." Unpublished.

Bailey, Martha J. 2010. "'Momma's Got the Pill': How Anthony Comstock and *Griswold v. Connecticut* Shaped US Childbearing." *American Economic Review* 100 (1): 98–129.

Bar, Michael, Moshe Hazan, Oksana Leukhina, David Weiss, and Hosny Zoabi. 2018. "Why Did Rich Families Increase Their Fertility? Inequality and Marketization of Child Care." *Journal of Economic Growth* 23 (4): 427–63.

Bar, Michael, and Oksana Leukhina. 2010. "Demographic Transition and Industrial Revolution: A Macroeconomic Investigation." *Review of Economic Dynamics* 13 (2): 424–51.

Barro, Robert J. and Jong Wha Lee. 2013. *Data for: "A New Data Set of Educational Attainment in the World, 1950–2010".* https://doi.org/10.1016/j.jdeveco.2012.10.001.

Baudin, Thomas. 2012. "The Optimal Trade-Off between Quality and Quantity with Unknown Number of Survivors." *Mathematical Population Studies* 19 (2): 94–113.

Beach, Brian, and W. Walker Hanlon. 2023. "Culture and the Historical Fertility Transition." *Review of Economic Studies* 90 (4): 1669–700.

Becker, Gary S., and H. Gregg Lewis. 1973. "On the Interaction between the Quantity and Quality of Children." *Journal of Political Economy* 81 (2): S279–88.

Becker, Sascha O., Francesco Cinnirella, and Ludger Woessmann. 2010. "The Trade-Off between Fertility and Education: Evidence from Before the Demographic Transition." *Journal of Economic Growth* 15 (3): 177–204.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Fortunata Makene, Linda Helgesson Sekei, Vincent Somville, and Bertil Tungodden. 2022. "On the Doorstep of Adulthood: Empowering Economic and Fertility Choices of Young Women." Unpublished.

Bergsvik, Janna, Agnes Fauske, and Rannveig Kaldager Hart. 2021. "Can Policies Stall the Fertility Fall? A Systematic Review of the (Quasi-) Experimental Literature." *Population and Development Review* 47 (4): 913–64.

Blanc, Guillaume. 2023. "The Cultural Origins of the Demographic Transition in France." Unpublished.

Bleakley, Hoyt, and Fabian Lange. 2009. "Chronic Disease Burden and the Interaction of Education, Fertility, and Growth." *Review of Economics and Statistics* 91 (1): 52–65.

Boldrin, Michele, Mariacristina De Nardi, and Larry E. Jones. 2015. "Fertility and Social Security." *Journal of Demographic Economics* 81 (3): 261–99.

Bongaarts, John, and Susan Cotts Watkins. 1996. "Social Interactions and Contemporary Fertility Transitions." *Population and Development Review* 22 (4): 639–82.

Caldwell, John C., and Pat Caldwell. 1987. "The Cultural Context of High Fertility in Sub-Saharan Africa." *Population and Development Review* 13 (3): 409–37.

Caucutt, Elizabeth M., Nezih Guner, and John Knowles. 2002. "Why Do Women Wait? Matching, Wage Inequality, and the Incentives for Fertility Delay." *Review of Economic Dynamics* 5 (4): 815–55.

Cervellati, Matteo, and Uwe Sunde. 2011. "Life Expectancy and Economic Growth: The Role of the Demographic Transition." *Journal of Economic Growth* 16 (2): 99–133.

Cinnirella, Francesco, Marc Klemp, and Jabob Weisdorf. 2017. "Malthus in the Bedroom: Birth Spacing as Birth Control in Pre-transition England." *Demography* 54 (2): 413–36.

Coale, Ansley J., and Susan Cotts Watkins, eds. 1986. *The Decline of Fertility in Europe: The Revised Proceedings of a Conference on the Princeton European Fertility Project.* Princeton University Press.

Collins, Matthew, Eleonora Guarnieri, and Helmut Rainer. 2025. "Fertility in Sub-Saharan Africa: The Interplay between Policy, Culture, and Intra-household Bargaining." CESifo Working Paper 11994.

d'Albis, Hippolyte, Paula E. Gobbi, and Angela Greulich. 2017. "Having a Second Child and Access to Childcare: Evidence from European Countries." *Journal of Demographic Economics* 83 (2): 177–210.

Danzer, Alexander M., and Lennard Zyska. 2023. "Pensions and Fertility: Microeconomic Evidence." *American Economic Journal: Economic Policy* 15 (2): 126–65.

Daudin, Guillaume, Raphaël Franck, and Hillel Rapoport. 2019. "Can Internal Migration Foster the Convergence in Regional Fertility Rates? Evidence from 19th Century France." *Economic Journal*

129 (620): 1618–92.
de la Croix, David, and Matthias Doepke. 2021. "A Soul's View of the Optimal Population Problem." *Mathematical Social Sciences* 112: 98–108.
de la Croix, David, and Fabio Mariani. 2015. "From Polygyny to Serial Monogamy: A Unified Theory of Marriage Institutions." *Review of Economic Studies* 82 (2): 565–607.
De Silva, Tiloka, and Silvana Tenreyro. 2017. "Population Control Policies and Fertility Convergence." *Journal of Economic Perspectives* 31 (4): 205–28.
Delventhal, Matthew J., Jesús Fernández-Villaverde, and Nezih Guner. 2024. "Demographic Transitions across Time and Space." NBER Working Paper 29480.
Desai, Jaikishan, and Alessandro Tarozzi. 2011. "Microcredit, Family Planning Programs, and Contraceptive Behavior: Evidence from a Field Experiment in Ethiopia." *Demography* 48 (2): 749–82.
D'Exelle, Ben, Aurélia Lépine, Richard Bakyono, and Ludovic D. G. Tapsoba. 2023. "Fertility and Polygyny: Experimental Evidence from Burkina Faso." *Journal of Development Economics* 164: 103134.
Doepke, Matthias. 2005. "Child Mortality and Fertility Decline: Does the Barro-Becker Model Fit the Facts?" *Journal of Population Economics* 18 (2): 337–66.
Doepke, Matthias, Anne Hannusch, Fabian Kindermann, and Michèle Tertilt. 2023. "The Economics of Fertility: A New Era." In *Handbook of the Economics of the Family*, Vol. 1, edited by Shelly Lundberg and Alessandra Voena, 151–254. North-Holland.
Doepke, Matthias, and Fabrizio Zilibotti. 2005. "The Macroeconomics of Child Labor Regulation." *American Economic Review* 95 (5): 1492–524.
Donald, Aletheia, Markus Goldstein, Tricia Koroknay-Palicz, and Mathilde Sage. 2024. "The Fertility Impacts of Development Programs." IZA Discussion Paper 17500.
Dupas, Pascaline, Camille Falezan, Marie Christelle Mabeu, and Pauline Rossi. 2023. "Long-Run Impacts of Forced Labor Migration on Fertility Behaviors: Evidence from Colonial West Africa." NBER Working Paper 31993.
Dupas, Pascaline, Seema Jayachandran, Adriana Lleras-Muney, and Pauline Rossi. 2025. "The Negligible Effect of Free Contraception on Fertility: Experimental Evidence from Burkina Faso." *American Economic Review* 115 (8): 2659–88.
Fernández, Raquel, and Alessandra Fogli. 2006. "Fertility: The Role of Culture and Family Experience." *Journal of the European Economic Association* 4 (2–3): 552–61.
Fernández, Raquel, and Alessandra Fogli. 2009. "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics* 1 (1): 146–77.
Fontenay, Sébastien, Paula Eugenia Gobbi, and Marc Goñi. 2025. "Fertility in Sub-Saharan Africa: The Role of Inheritance." CEPR Discussion Paper 20275.
Galor, Oded. 2012. "The Demographic Transition: Causes and Consequences." *Cliometrica* 6 (1): 1–28.
Galor, Oded, and David N. Weil. 1996. "The Gender Gap, Fertility, and Growth." *American Economic Review* 86 (3): 374–87.
Galor, Oded, and David N. Weil. 2000. "Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond." *American Economic Review* 90 (4): 806–28.
Gay, Victor, Paula Eugenia Gobbi, and Marc Goñi. Forthcoming. "Revolutionary Transition: Inheritance Change and Fertility Decline." *Journal of Political Economy*.
Giuliano, Paola and Nathan Nunn. 2018. *Data for: "Ancestral Characteristics of Modern Populations."* https://doi.org/10.1080/20780389.2018.1435267.
Gobbi, Paula E., and Marc Goñi. 2021. "Childless Aristocrats: Inheritance and the Extensive Margin of Fertility." *Economic Journal* 131 (637): 2089–118.
Gobbi, Paula E., Anne Hannusch, and Pauline Rossi. 2026. *Data and Code for: "Family Institutions and the Global Fertility Transition."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E239791V1.
Goldin, Claudia. 2025. "Babies and the Macroeconomy." *Economica* 92 (367): 675–700.
Golosov, Mikhail, Larry E. Jones, and Michèle Tertilt. 2007. "Efficiency with Endogenous Population Growth." *Econometrica* 75 (4): 1039–71.
Guinnane, Timothy W. 2011. "The Historical Fertility Transition: A Guide for Economists." *Journal of Economic Literature* 49 (3): 589–614.
Guner, Nezih, Ezgi Kaya, and Virginia Sánchez-Marcos. 2024. "Labor Market Institutions and Fertility." *International Economic Review* 65 (3): 1551–87.
Haddad, Joanne. 2023. "Settlers and Norms." Unpublished.

Hajnal, John. 2017. "European Marriage Patterns in Perspective." In *Population in History: Essays in Historical Demography*, edited by D. V. Glass and D. E. C. Eversley, 101–44. Routledge.

Hartmann, Betsy. 1997. "Population Control I: Birth of an Ideology." *International Journal of Social Determinants of Health and Health Services* 27 (3): 523–40.

Hazan, Moshe, David Weiss, and Hosny Zoabi. 2021. "Marketization and the Fertility of Highly Educated Women Along the Extensive and Intensive Margins." CEPR Discussion Paper 16647.

Hazan, Moshe, David Weiss, and Hosny Zoabi. 2023. "Highly Educated Women Are No Longer Childless: The Role of Marketization." *Economics Letters* 230: 111226.

Herzer, Dierk, Holger Strulik, and Sebastian Vollmer. 2012. "The Long-Run Determinants of Fertility: One Century of Demographic Change 1900–1999". *Journal of Economic Growth* 17 (4): 357–85.

Jaeger, David A., Theodore J. Joyce, and Robert Kaestner. 2020. "A Cautionary Tale of Evaluating Identifying Assumptions: Did Reality TV Really Cause a Decline in Teenage Childbearing?" *Journal of Business and Economic Statistics* 38 (2): 317–26.

Jones, Eric. 2003. *The European Miracle: Environments, Economies and Geopolitics in the History of Europe and Asia*. 3rd ed. Cambridge University Press.

Kearney, Melissa S., and Phillip B. Levine. 2015. "Media Influences on Social Outcomes: The Impact of MTV's *16 and Pregnant* on Teen Childbearing." *American Economic Review* 105 (12): 3597–632.

Kearney, Melissa Schettini, and Phillip B. Levine. 2025. "Why Is Fertility So Low in High Income Countries?" NBER Working Paper 33989.

Kim, Seongeun, Michéle Tertilt, and Minchul Yum. 2024. "Status Externalities in Education and Low Birth Rates in Korea." *American Economic Review* 114 (6): 1576–611.

La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. "Soap Operas and Fertility: Evidence from Brazil." *American Economic Journal: Applied Economics* 4 (4): 1–31.

Lambert, Sylvie, and Pauline Rossi. 2016. "Sons as Widowhood Insurance: Evidence from Senegal." *Journal of Development Economics* 120: 113–27.

Lesthaeghe, Ron J. 1977. *The Decline of Belgian Fertility, 1800–1970*. Princeton University Press.

Lesthaeghe, Ron J. 1989. *Reproduction and Social Organization in Sub-Saharan Africa*. University of California Press.

Mahler, Lukas, Michèle Tertilt, and Minchul Yum. Forthcoming. "Policy Concerns in an Era of Low Fertility: The Role of Social Comparisons and Intensive Parenting." *Brookings Papers on Economic Activity*.

Melki, Mickael, Hillel Rapoport, Enrico Spolaore, and Romain Wacziarg. 2024. "Cultural Remittances and Modern Fertility." NBER Working Paper 32990.

Miller, Grant, and Kimberly Singer Babiarz. 2016. "Family Planning Program Effects: Evidence from Microdata." *Population and Development Review* 42 (1): 7–26.

Murdock, George Peter. 1967. *Data for: "Ethnographic Atlas: A Summary."* https://doi.org/10.2307/3772751.

Murphy, Tommy E. 2015. "Old Habits Die Hard (Sometimes): Can *département* Heterogeneity Tell Us Something about the French Fertility Decline." *Journal of Economic Growth* 20 (2): 177–222.

Notestein, Frank Wallace. 1952. *Economic Problems of Population Change*. Oxford University Press.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford University Press.

Perrin, Faustine. 2022. "On the Origins of the Demographic Transition: Rethinking the European Marriage Pattern." *Cliometrica* 16 (3): 431–75.

Preston, Samuel H. 1978. *The Effects of Infant and Child Mortality on Fertility*. Academic Press.

Pritchett, Lant H. 1994. "Desired Fertility and the Impact of Population Policies." World Bank Policy Research Working Paper 1273.

Rossi, Pauline. 2019. "Strategic Choices in Polygamous Households: Theory and Evidence from Senegal." *Review of Economic Studies* 86 (3): 1332–70.

Rossi, Pauline, and Mathilde Godard. 2022. "The Old-Age Security Motive for Fertility: Evidence from the Extension of Social Pensions in Namibia." *American Economic Journal: Economic Policy* 14 (4): 488–518.

Rossi, Pauline, and Yun Xiao. 2024. "Spillovers in Childbearing Decisions and Fertility Transitions: Evidence from China." *Journal of the European Economic Association* 22 (1): 161–99.

Rossi, Pauline, and Yun Xiao. 2026. "Left Over or Opting Out? Squeeze, Mismatch and Surplus in Chinese Marriage Markets." *Journal of Development Economics* 179: 103629.

Sage, Mathilde. 2025. "Children Are a Poor Women's Wealth: How Inheritance Rights Affect Fertility." Louvain Institute of Data Analysis and Modeling in Economics and Statistics Discussion Paper 2025/04.

Spolaore, Enrico, and Romain Wacziarg. 2022. "Fertility and Modernity." *Economic Journal* 132 (642): 796–833.

Strulik, Holger. 2004. "Child Mortality, Child Labour and Economic Development." *Economic Journal* 114 (497): 547–68.

Strulik, Holger. 2008. "Geography, Health, and the Pace of Demo-economic Development." *Journal of Development Economics* 86 (1): 61–75.

Tertilt, Michéle. 2005. "Polygyny, Fertility, and Savings." *Journal of Political Economy* 113 (6): 1341–71.

Todd, Emmanuel. 1984. *L'enfance du Monde: Structures Familiales et Développement*. Seuil.

United Nations. 2020. *World Fertility and Family Planning 2020: Highlights*. UN Department of Economic and Social Affairs, Population Division.

Vogl, Tom. 2025. "Fertility Decline and Educational Progress among African Women and Children." *Demography* 62 (5): 1581–1606.

Voigtländer, Nico, and Hans-Joachim Voth. 2013. "How the West 'Invented' Fertility Restriction." *American Economic Review* 103 (6): 2227–64.

Wolpin, Kenneth I. 1997. "Determinants and Consequences of the Mortality and Health of Infants and Children." In *Handbook of Population and Family Economics*, Vol. 1A, edited by Mark R. Rosenzweig and Oded Stark, 483–557. Elsevier.

World Bank. 2025. *Data for: "World Development Indicators."* World Bank. https://databank.worldbank.org/source/world-development-indicators (accessed June–July 2025).

Journal of Economic Perspectives—Volume 40, Number 1—Winter 2026—Pages 71–92

Global Labor Mobility between Shrinking and Growing Labor Forces

Lant Pritchett

Almost a half-century ago, Robert McNamara (1977), then president of the World Bank, stoked "population bomb" fears saying: "Short of thermonuclear war itself, population growth is the gravest issue the world faces over the decades immediately ahead." With the total fertility rate in most countries already below, or rapidly nearing, replacement, such predictions seem quaint. Perhaps the new threat to humankind's very long-run prosperity (Jones 2022) and even very long-run survival could be a depopulation bomb (Eberstadt 2022; Spears and Geruso 2025). But the most pressing "new population bomb" (Goldstone 2010), one coming with near certainty in the medium-run (defined, per Keynes, as the period over which the median reader is not dead), is not total global population but aging in the richest countries and youth bulges in some developing country regions, which dramatically shifts the global composition of the labor force aged population.

The total fertility rate fell below 2.1 (a rough and ready estimate of the replacement rate) in the rich countries before 1980 and has fallen even further. As a consequence of this and extensions of lifespan, the nonmigrant working-aged population is headed for a substantial decline, and yet the number of elderly citizens will continue to rise. This same aging will affect, even more sharply, China. In contrast, most low-income countries have experienced fertility declines recently and are now, or are near to, reaching replacement levels in fertility. Momentum will lead to rising working-aged populations and rapid aging. Africa is distinctive in that

■ *Lant Pritchett is Visiting Professor of Practice, London School of Economics, School of Public Policy, London, United Kingdom. His email address is lant.hks@gmail.com.* *Due to the data editor's inability to find a replicator with access to the software used, the results of this paper were not reproduced. However, the data editor reviewed all instructions and data files, and found them to be in compliance with AEA's Data and Code Availability Policy.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251461.

total fertility rates are still well above replacement and hence the young and labor force aged populations will rise sharply to 2050.

With working-aged population falling in high-labor-productivity places and rising in low-productivity places, the potential gains from allowing workers to move from low-productivity, youth-bulge regions to high-productivity, labor-scarcity regions are in the trillions of dollars (Clemens 2011). Over the next few decades, relaxing the current constraints on this movement is by far the largest known policy-based economic opportunity for improving global human well-being.[1] But this potential economic gain is unlikely to become a reality unless the politics in the rich world can adopt pathways for "rotational" labor migration that are potentially more politically palatable than permanent migration.

Fertility and Labor Force, by Region

Looking ahead to 2050, in one group of countries that includes all of the world's richest economies, China, and (most of the) former Soviet Union, the *absolute* number of youth is going to fall and the *absolute* number of those over 65 is going to rise, and hence population growth rate is going to gradually fall and become negative. In the absence of migration, these changes will push all these countries into demographic *terra incognita*. No human society has even seen the degree of inverted demographic pyramids, and the accompanying high ratios of those 65+ to labor force aged, that these countries will experience. In the second group, which includes most of the rest of the world less Africa, the falls in fertility are more recent and hence momentum will cause absolute increases in youth, but at a slowing pace, along with large increases in those 65+. In Africa, the fall in fertility has come later and has not been as rapid. Hence, in these countries the absolute number of those of labor force age is going to more than double by 2050—and yet there will be still be population aging as those 65+ will triple in Africa. In this section, I explore the dynamics and determinants of these population shifts.

The Fertility Decline, by Region

There are three major sources of country specific data on recent past and projected fertility. The UN Population Division produces World Population Projections, which, for all their past faults, have been the standard source (UN 2024). The Institute for Applied Systems Analysis (IIASA), as part of their Shared Socioeconomic Pathways (SSP) scenarios, produces forecasts of fertility and population by age and by level of schooling (K. C. et al. 2024). The Global Burden of Disease (GBD) project, as part of its effort to estimate future health conditions, produces estimates of fertility and population by age (GBD 2024). As they all rely on roughly

[1] "Known" is an important quality. If (inclusive) economic growth was sustained at a much higher rate, this would produce similar gains in human wellbeing, but that is comparing an outcome, where the actions to reliably achieve that outcome are far from certain, to a single discrete policy shift.

the same original country sources, the cross-national correlations of their estimates of past changes are very high, and medium-run projections are quite high as well. Only the very long-run fertility predictions differ substantially: the IIASA has nearly all countries converging on a similar total fertility rate by 2100, which produces a weak correlation with either the UN or GBD projections.

Using the simple average total fertility rate from the three sources for each country, Figure 1 shows the annual population weighted total fertility rate for ten geographic regions from 1950 to 2025, with the embedded tables showing average fertility forecasts for 2050 and 2100.

Panel A of Figure 1 shows the five regions where total fertility rate fell below 2.1 before 1991. Three of those regions can be categorized as advanced, democratic, and rich industrial countries: *Asia and Pacific* (Japan, Korea, Australia, New Zealand), *Europe* (26 countries), and *North America* (United States and Canada). Adding "democratic" and "industrial" countries to high GDP per capita divides the historical "developed" economies from the oil-rich Gulf states and Singapore, which have quite distinct demographic patterns and much higher labor mobility. *China* is its own region. The fifth region is an amalgam of the former Soviet Union and Eastern Europe along with Mongolia and North Korea.

Panel B of Figure 1 shows the five regions which transitioned to lower fertility more recently: *Latin America and the Caribbean* (LAC), *South Asia*, *South East Asia and Pacific* (excluding the four Asia and Pacific countries included above), *Africa* (both sub-Saharan and North, hence including Morocco to Egypt), and the region the United Nations refers to as *West Asia* and the World Bank and others call the Middle East, which includes Israel, Syria, and Lebanon as well as Turkey and Iran.

Panels A and B of Figure 1 jointly illustrate four facts about the recent history and projections of fertility.

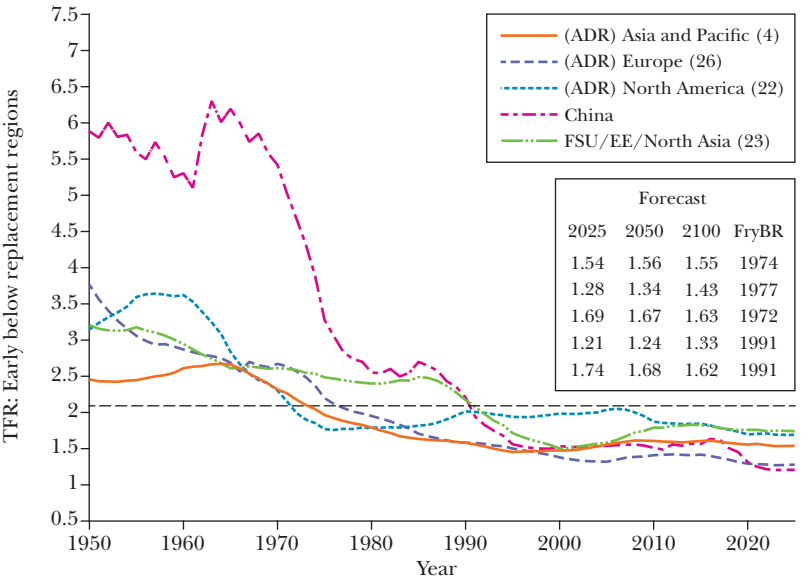
First, the total fertility rate (TFR) in all three advanced, democratic, and industrialized regions (and in each of the 32 countries individually) fell below 2.1 before 1980 (more than 45 years ago). TFR has not rebounded significantly from its lowest point in any of these countries: all are still well below replacement.[2] Moreover, the projections (shown in the embedded tables in Figure 1) show these regions remaining at least half a birth per woman below 2.1 in 2100. Projections 75 years into the future necessarily have large uncertainty, but simple extrapolations of the recent past, where TFR has been low and stable or declining, and the methods used in all three sources (the UN Population Division, the Institute for Applied Systems Analysis, and the Global Burden of Disease) tell the same basic story.

Second, China is an important and distinctive story: important, because it is a population behemoth with more than one-sixth of global population; distinctive,

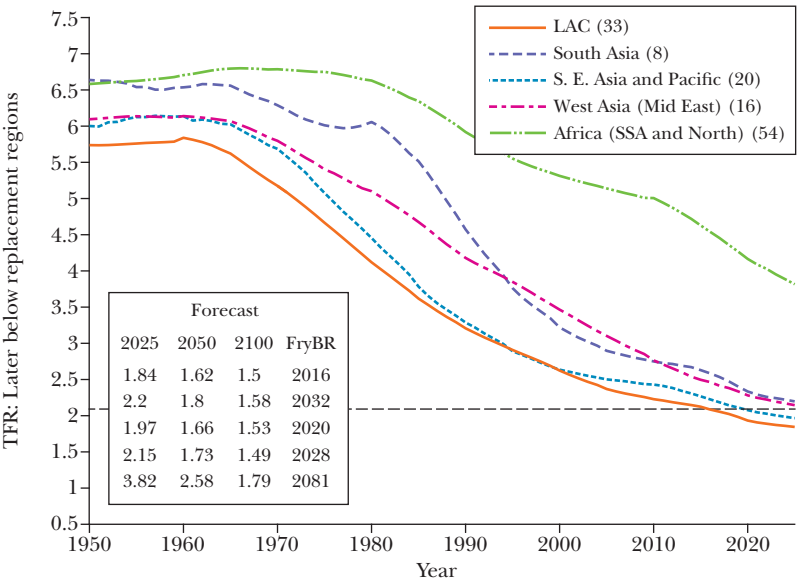
[2]Goldin (2025) analyzes a phenomenon in which the advanced, democratic, and rich industrial countries that have the "lowest low" fertility (like Spain, Italy, Japan, and Korea) had a later onset of rapid growth than did the "early below replacement" countries (like France, Germany, and Sweden), and hence fell faster to a lower level. In the next section, we show that this makes a difference to the magnitude of the coming aging.

Figure 1
The Timing and Depth of Fall in Fertility Differs Widely Across Regions of the World

Panel A. Trajectory of total fertility rate in early transition regions



Panel B. Trajectory of total fertility rate in later transition regions



Source: Author's calculation of population weighted average total fertility rate from data/projections of UNPD, GDB, and IIASA.
Note: In panel A, ADR stands for "advanced, democratic, rich." In the embedded tables, the numbers refer to the number of countries in each region. The final column of the table, FyrBR, stands for first year below replacement rate.

because of its recent rapid trajectory from very high fertility (a total fertility rate of 5.42 as late as 1970) to very low (a TFR of 1.21 in 2025). The common phrase that "China will get old before it gets rich" is the result of hitting replacement fertility at levels of income much lower than the advanced, democratic, and rich industrial countries: China's GDP per capita in the first year it hit TFR of 2.1 was only $2,600 (using a purchasing power parity exchange rate). Also distinctive are the waves in China's decline in TFR: a very rapid fall from 1970s to 1980s, a plateau in the 1980s, another decline, another plateau from late 1990s to late 2010s, and then another decline. The speed of China's fertility decline makes its aging up to 2050 similarly much more rapid.

Third, outside of Africa, all regions (and nearly all countries in those regions) have both experienced large declines and are at or nearing a total fertility rate of 2.1. This fall in fertility does not appear to be tapering off to stop at 2.1. The similarity in the pace in TFR decline and the convergence to similar, near replacement, levels by 2025 in these four developing country regions is striking.

Fourth, Africa's fertility is distinctive: its total fertility decline started later (around 1980 versus 1960s for other regions), from higher levels (around 6.5 in 1980 by which times other regions were much lower), and was more gradual. Hence Africa's total fertility rate in 2025 is 3.82, well above the next highest region (South Asia at 2.2), and is projected to be at 2.58, still well above replacement, in 2050. Not until 2081 is Africa as a region forecast to reach replacement levels: around 100 years later than any of the regions of advanced, democratic, and rich industrial countries, 90 years later than China, and 50 years later than any other developing country region (South Asia is projected to reach 2.1 in 2032).

Why Fertility Is Unlikely to Rebound

Understanding fertility decline is, to first order, understanding why women (and men) want fewer children. I make no attempt here to review the massive literature, both theoretical and empirical, on correlates and causes of fertility and its decline. Here, I just touch on five themes to emphasize that the global secular decline in fertility (and its regional pattern) is unlikely to stop or reverse, regardless of whether governments adopt pro-fertility policies.

First, it is widely accepted that country-level secular declines in fertility are almost entirely "accounted for" in a correlation sense by changes in GDP per capita, female schooling, and child mortality, although sub-Saharan Africa's slow fertility transition appears to have followed a moderately different pattern of economic structural transformation (Büttner et al. 2024).[3] This is one of those relatively

[3]More formally, a regression of average total fertility rate across all countries for all five-year intervals from 1950 to 2020 on just four indicators (GDP per capita, female years of schooling, and under five mortality rate) using a modestly flexible functional form—cubic in GDP per capita (Feenstra, Inklaar, and Timmer 2015); quadratic in female years of schooling (Barro and Lee 2021) and the under 5 mortality rate (UNICEF 2025), plus a sub-Saharan Africa dummy produces an R^2 of .854. For regression results and an illustrative figure, see Pritchett (2026).

rare cases where the time series mostly tracks the cross-section as, controlling for progress in these three correlates, there is no secular shift in the total fertility rate.

Second, these three strong cross-country correlates of fertility are widely regarded as normatively good things. In fact, measures of per capita GDP, schooling, and health were the *only* elements of the original UN Human Development Index. No one advocates reducing falls in fertility by stopping progress in education, child health, or economic prosperity.

Third, for many years the estimates of future fertility assumed that fertility higher or lower than 2.1 was on a long-run glide path to the replacement. This was an assumption of convenience to avoid the knife-edged properties of long-run demographic projections, as there never has been any descriptive behavioral model—from any discipline—suggesting that fertility at replacement rates was an equilibrium of individual's choices (Pritchett and Viarengo 2012). Weil's (2024) title summarizes the point succinctly: "Replacement fertility is neither natural nor optimal nor likely." As the assumed recovery of total fertility rates in rich countries upward towards replacement failed to happen, the UN projections have now given up on forcing this convergence to 2.1 into their projections (Weil 2024).

Fourth, using the Demographic and Health Survey data from the World Bank, the case can be made that nearly all differences in child-bearing both across countries in levels and over time are associated with the level and changes in how many children women report they want to have (Pritchett 1994; Gunther and Harttgen 2016, with some caveats about fertility in sub-Saharan Africa). This claim is not uncontroversial, because some have argued that women would prefer fewer children but have an "unmet need" for contraception.[4] My own belief is that the existing measures of "unmet need" actually almost exclusively reflect "demand side" reasons women have for not using contraception rather than "supply side" or "lack of access" constraints (Senderowicz and Maloney 2022). For example, a recent large randomized control trial of providing individuals with vouchers covering 100 percent of contraceptive costs in a high-fertility country, Burkina Faso, found zero impact (with small enough standard errors even modest effects could be ruled out) on live births, pregnancies, or even the use of medical contraception over three years—and, strikingly, this zero impact was found even among women measured as having "unmet need" at the baseline (Dupas et al. 2025).

Taking these factors together, government attempts to raise fertility seem unlikely to have a substantial effect. The "all in" costs and benefits of having a child—monetary, temporal, psychic, social—are very large. Hence, it would be very surprising if there were cheap and effective ways for governments to raise fertility conditional on existing socioeconomic realities individuals and couples

[4]This controversy continues. The Global Burden of Disease (2024) scenarios for future fertility assume that "unmet need satisfied" is a *causal* policy driver of fertility. That there is a relatively strong empirical correlation (at country and individual level) between contraceptive use and fertility is not disputed. The controversy is over whether this correlation represents a causal relationship from contraceptive "access" (low cost, in a broad definition of cost) to reduced births or that people who want fewer children are more likely to use contraception.

face (*The Economist* 2024). The Global Burden of Disease (2024) projections assume (based on the review of the empirical literature about impacts in richer countries of pro-natal policies by Bergsvik, Fauske, and Hart 2021) that the adoption of the known effective pro-natal policies (at feasible levels) would raise the total fertility rate in 2050 by .2, which would, for instance, raise their projected total fertility rate average in the advanced and democratic economies of Europe from 1.34 to 1.54.

Moreover, even if there are policies to change future fertility, these will make very little difference in the next few decades. The demography of the labor force aged in 2050 is pretty "baked in" (Goldstone forthcoming), as everyone who is going to be in the "prime" labor force age bracket of 25 to 49 in 2050 has to be already alive in 2026.

Growth of the Labor Force Aged (15–64) versus the Elderly (65+)

A global decline in fertility, occurring at different times across regions, leads to differential growth in the size of the labor force aged population (ages 15–64) and the elderly population age 65 and over. The International Institute for Applied Systems Analysis (IIASA) produced estimates of world population by age, sex, and educational status for 201 countries with 2020 as the base year and projections to 2050 and 2100 (K. C. et al. 2024). Here, I use their "Middle of-the-Road/ Continuation" scenario with zero migration (SSP2-ZM) but the results are broadly similar using the UN Zero Migration scenario. This zero migration scenario is not a "projection"—there is and will be some level of migration to 2050—but is used to illustrate cleanly the demographic implications of natural increase from births and deaths alone. I calculate the aged 15–64 and the over-65 population for each country.

Table 1 illustrates three key features about the world's medium-term demographic future to 2050, using the ten regions that appeared in Figure 1.

First, the primary demographic challenge to the year 2050 is not rising or declining total population but a shift in the age distribution of the population. For the advanced, democratic, rich, and industrial regions (first three rows) the labor force aged population falls by 121 million, an 18 percent fall, while people 65+ rise by 111 million, a 54 percent increase. In China, the 15–64 population falls by 211 million (20 percent), while its 65+ population rises by 210 million (119 percent). These changes are substantial to dramatic for each age group, but largely offsetting for the total population.

In the early fertility regions, the ratio of working-aged to population over 65 falls substantially from 2020 to 2050 to historically unprecedented levels (Goldstone forthcoming). In 2020 Japan had only 1.98 people aged 15–64 for each person over 65, which is the lowest observed in any country, ever. By 2050 in the advanced economies taken together, this ratio drops from 3.5 to 1.8: the typical advanced country will be at levels lower than seen in human history to date.

In China, the ratio of working-aged to population over 65 falls spectacularly from 5.3 in 2020 to 1.5 in 2050. Within the advanced, democratic, and rich countries are some with "lowest low" fertility (Kohler, Billari, and Ortega 2002; Goldin 2025),

Table 1
By 2050, Nearly All Labor Force Aged Expansion Happens in Africa and South Asia

Region (number of countries)	2020 15–64	2020 65+	2050 15–64	2050 65+	Change 2020 to 2050: 15–64 Absolute	15–64 %	65+ Absolute	65+ %	Region as percent of net increase in pop. 15–64
Early Transition to Lower Fertility Regions									
Advanced, democratic, rich, industrial countries (32)	667	204	547	315	−121	−18.1%	111	54.4%	−12.0%
Europe (26)	292	94	231	142	−61	−20.8%	48	−51.6%	−6.0%
Japan, Korea, Australia, New Zealand (4)	131	50	94	69	−37	−28.6%	20	−39.4%	−3.7%
N. America (2)	244	60	222	103	−23	−9.3%	43	−71.3%	−2.3%
China (1)	996	177	784	387	−211	−21.2%	210	118.9%	−21.1%
Former USSR (23)	272	56	248	90	−23	−8.6%	34	61.3%	−2.3%
Later Transition to Fertility Regions									
Latin America (38)	437	57	490	144	53	12.1%	88	155.5%	5.3%
W. Asia (Middle East) (16)	235	20	306	77	71	30.0%	58	293.7%	7.0%
South East Asia and Pacific (22)	458	49	524	135	66	14.3%	86	172.9%	6.6%
South Asia (8)	1,230	115	1,600	303	370	30.1%	187	162.9%	37.0%
Africa (57)	751	46	1,549	140	798	106.2%	94	201.6%	79.6%
Total	5,046	724	6,048	1,591	1,003	19.90%	979	105.5%	100.0%

Source: Author's calculations with IIASA data (K. C. et al. 2024) downloaded from Wittgenstein Centre Human Capital Explorer.

and by 2050 this very low fertility leads to ratios of labor force aged to over-65 population even lower than China's projection: Italy to 1.25, Korea to 1.25, Spain to 1.26, and Greece to 1.43.

Second, while the absolute labor force aged population falls in the early fertility regions, more than 100 percent of the *net* increase in the labor force aged population happens in just two regions: South Asia and Africa. From 2020 to 2050, the global net population aged 15–64 rises by one billion. The five early fertility decline regions will see their labor force aged fall by a combined 355 million. The three regions of Latin America, West Asia/Middle East, and South East Asia and Pacific (76 total countries) have a modest combined rise in the aged 15–64 population of 190 million. However, while the fertility evolution in South Asia is similar to these countries, by dint of its massive existing population the workforce-aged population of South Asia grows by 370 million. Africa, the slow fertility transition region (Figure 1, panel B), will see its working-age population growing by 798 million from 2020 to 2050. Africa alone will account for nearly 80 percent of the growth in the world's labor force aged to 2050. Africa and South Asia combined account for a 1.17 billion increase, well more than 100 percent of the net global growth.

Third, the ratio of working-aged to elderly population is falling everywhere, even in regions with continued growth in population and growing workforce-aged. For example, the working-aged population in South Asia grows by 370 million from 2020 to 2050, a 30 percent rise, but that same region will see its 65 and over

population rise from 115 million to 305 million, a near tripling. In the Africa region, the working-aged population doubles from 2020 to 2050, but the elderly population triples.

How Many More Working-Aged People to Maintain the Dependency Ratio?

Let us presume that the projections for the size of the over-65 elderly population (without migration) in 2050 hold true. How many more people would it take to keep the ratio of working-aged to elderly somewhat stable? There are two ways to think about this question: looking at the total working-aged population divided by the number of elderly, or looking at the number of workers—that is, the total working-aged population multiplied by the labor force participation rate and then divided by the number of elderly.

The answers have practical and policy implications. For public finance, in most countries the public pensions for the elderly are financed through taxes on current workers. Thus, when the ratio of workers/elderly falls, the public retirement system comes under stress. More broadly, support for the elderly happens not just though public pensions, but also through support from younger family, friends, and volunteers. A decline in the ratio of the working-aged population (whether actually holding jobs or not) to the elderly makes these nonmonetary support systems harder to sustain as well.

The over-65 population of the advanced, democratic, rich, and industrialized countries (shown in the first three rows of Table 1) is forecast to increase by 111 million from 2020 to 2050. For these countries, the ratio of the working-aged to the elderly population in 2020 for the median country was 3.3. To keep that ratio constant, these countries would need the working-aged population to increase by about 363 million (that is, 3.3 × 111 million). But the workforce-aged population is not going to grow—it is going to shrink by 121 million—so the gap between actual and a constant ratio labor force is 484 million people.[5] To put that number in perspective, the zero-migration projections for the total working-aged population in these advanced countries in 2050 is only 547 million.

Using labor force participation rates by five-year age cohorts and sex, one can convert the purely demographic projections of ratios of age groups into a number of workers and elderly. Following the approach in Pritchett (2024), I work though this exercise for the 32 advanced, democratic, and rich industrial countries using labor force participation data by sex and five-year age groups in 2020 from the International Labour Organization (ILOSTAT 2024).[6]

[5]The formula for the demographic gap as a function of a target ratio is straightforward:

$$Demographic\ Gap\big(Target\ ratio\big)_{2050} = \Big(Target\frac{LFA}{65+}\ ratio\Big)_{2050} \times pop'l^{65+}_{2050} - pop'l^{LFA}_{2050}.$$

[6]A primary challenge with implementing this approach that the "top code" of ILO-reported labor force participation rates (LFPR) is for the 65 and over group. But of course, there will be a substantial difference between, say, the LFPR in the 65–70 age group as compared to that of the 85–90 age group. I handle this by assuming that the LFPR for each five-year cohort over is half that of the just younger age group (for example, 80–84 is half that of 75–79) to create a parametric estimate for each category out to

In the advanced European countries, the over-65 population is projected to grow by 43.3 million from 2020 to 2050. Their current ratio of workers/elderly is 2.4. Thus, to maintain this ratio by 2050 would require roughly $2.4 \times 43.3 = 104$ million *more* workers. But in the no-migration and constant labor force participation scenario, these countries will actually have 45 million *fewer* workers in 2050. The 2050 population weighted ratio of workers to elderly falls to 1.4 for advanced Europe as a whole, and the ratio in *every* European country—except Sweden at 1.9—is below Japan's 2020 ratio of 1.83. Indeed, in this scenario there are about one or fewer workers for every person over 65 in Spain (1.00), Greece (1.04), and Italy (0.88).

The labor force and aging dynamics of the United States and Canada are quite different from Europe. Their 2020 ratio of labor force to over-65 population is closer to 3. Also, the projected fall in the labor force in the United States from 2020 to 2050 is proportionately much smaller. For example, the United States over this time has 8 million fewer workers from a base of 220 million 2020, a 6 percent fall, whereas Italy has a decline of 7.7 million workers from the smaller base of 38 million in 2020, a 32 percent fall. In the scenario with no immigration and a constant labor force participation rate, the ratio of labor force to over-65 population falls to 1.86 by 2050. The number of additional workers needed depends critically on the ratio sought for 2050. Maintaining the 2020 ratio in North America requires a very large number of additional workers, about 123 million. But falling only to a ratio of 2.11 (the 2020 10th percentile of the rich countries) in 2050 would require only 31 million additional workers.

Japan has been the canary in the coal mine of low fertility and an aging economy and society. Japan's ratio of labor force to elderly had already fallen to 1.83 in 2020, and the future projected decline to 1.24 by 2050 is thus relatively modest, a fall of .59 compared to a projected fall of 1.0 in average in Europe (2.4 to 1.4) and 1.3 (3.0 to 1.7) in North America. For Japan to have a ratio of labor force to elderly of 2.11 by 2050 would require 33 million more workers by 2050.

South Korea is an extreme example of the demographic dynamics of aging. In 2023, Korea's total fertility rate declined to 0.72, with deaths exceeding births by over 100,000. In the projections, Korea's ratio of labor force to elderly falls from 3.46 in 2020 to just 1.16 in by 2050. Hence, the 17.1 million additional workers needed to reach a ratio of 2.11 would be 45 percent of its projected 2050 labor force.

A country seeking to maintain its 2020 ratio of labor force to elderly, or even just to raise those ratios substantially from the projections for 2050, faces a difficult task if it limits itself to domestic policies. It takes 25 years to produce a 25-year-old. Hence, even if the most optimistic changes in fertility took place immediately, these would have little effect on the ratio of the labor force aged population by 2050.

100+, and then adjusting the level of those LFPRs to match the 2020 reported aggregate 65+ LFPR. My base case calculation of the 2050 labor force uses the estimates of the 2050 population age/sex structure from the 2024 UN World Population Prospects Zero Migration scenario and each country's 2020 LFPRs by five-year age groups and sex.

Moreover, there is only limited scope for many countries to exceed their projected labor force for 2050 by raising the labor force participation rates. For prime-aged workers, labor force participation rates are already fairly high. For example, across the 32 advanced, democratic, rich, and industrialized countries, the 2020 labor force participation rate for people aged 40–44 is 93 percent for males and 84 percent for females. Pritchett (2024) shows the needed additional workers across a variety of assumptions about 2050 labor force participation. Germany's projected labor force gap falls from 31.5 million with constant 2020 labor force participation to only 16 million in the "high" scenario for 2050 labor force participation. However, in this "high" scenario, the labor force participation of those over 65 would need to rise in Germany from 21 percent to 72 percent, which seems a social and political impossibility.

Cross-Border Mobility: Economics and Politics

While there has been a global decline in fertility and a resulting shift toward the elderly across the global population, there is not a global scarcity of labor. Rather, the regional differences in the timing and magnitude of fertility imply that in the next few decades there will be increasingly severe scarcity of young workers in some regions (the rich countries and China) and a youth bulge abundance in others (Africa and South Asia). The challenge of the birth-dearth-induced labor scarcity in the rich world cannot be wholly solved by raising fertility (too late) or increased labor force participation (too little). Policies that allow for movement of working-aged people from labor-abundant, low-productivity places to labor-scarce, high-productivity places is an obvious economic win-win, but sufficient expansion of just existing pathways seems a political nonstarter. New modes of movement are needed.

Many dismiss the idea of substantially *more* movement of people as a response to the coming shifts in the world's demography as not worth discussing, as substantial increases seem so far outside the realm of what is currently politically possible. But the changes in age structure of national populations that will be arriving in the next few decades will be so large that I expect them to shift the boundaries of the politically possible. What will countries be willing to attempt when they face the reality of trying to fund public pensions with a ratio of roughly one worker for every retiree? In particular, I argue that a substantial expansion in legal pathways for time-limited labor mobility can create massive win-win gains for both labor-scarce, high-productivity countries and youth-abundant, low-productivity countries and, if done well, be politically palatable.

Economics of Movement

The simplest parts of the economics of cross-border movement from poor to rich countries are the wage gains to movers and hence the supply of potential movers. A variety of researchers have found that migrants from lower- to

higher-income economies experience large real wage gains, with the ratio of wages for an equivalent low- to moderate-skill worker varying across countries but from three to ten, with a modal value around five.[7] An obvious concern in such studies is how to adjust for potential selection effects and not have average estimates for "migrants" be affected by highly selected "superstar" movers. For workers in "core skill" occupations and/or modest levels of schooling (below tertiary), the selection-corrected results are robust across a wide variety of methods: randomized control trials (McKenzie, Stillman, and Gibson 2010); natural experiments due to different selection mechanisms that affect the possibility of migration (Mobarak, Sharif, and Shrestha 2023; Clemens 2011); econometric methods to estimate bounds on selection effects (Clemens, Montenegro, and Pritchett 2019, relying on Oster 2019 adjustments); wages for the same jobs in the same firms but in different countries (Ashenfelter and Jurajda 2001); and wage differences of workers across countries in the same detailed low-skill occupations (Pritchett and Hani 2020). These micro-economic estimates are consistent with (nearly all) aggregate estimates of total factor productivity differences across countries and their estimates of wage differentials between high- and low-income countries for human capital equivalent labor (Pritchett and Hani 2020).

Perhaps unsurprisingly given these potential wage gains, the supply of people willing to move from low- to high-productivity to work, either permanently or for just a period, is large. As part of its international polling, Gallup asked people: "Ideally, if you had the opportunity, would you like to move permanently to another country, or would you prefer to continue living in this country?" Then, of those who express a desire to move, "To which country would you like to move?" Gallup calculated the youth population in each country if everyone's migration preferences were met. The median expansion of the youth population in the countries in the top one-third of labor compensation per hour was 111 percent (a rough doubling), with the United States at almost exactly the median and some smaller, attractive destinations (like Canada, Switzerland, Australia) more than tripling their youth populations if all respondents moved to their ideal location (Pritchett and Hani 2020). The willingness of workers to move across borders when the wage gains are much smaller, such as inter-regional movements within Africa or Latin America or movements to the Gulf region, is also consistent with an enormous willingness to move at the existing wage differentials between potential low- and high-productivity regions and countries.

Those who would be willing to move from low- to high-productivity countries face border controls and legal constraints. Every high-productivity country imposes and enforces legal restrictions on who can live and work in their country. One gruesome sign of the constraints on the widespread desire to work in a

[7]All of these reported wage differentials fully adjust for purchasing power differences (that prices are higher in rich countries). Therefore, if movers spend only some of the wage gains in the host country and some at the source country (via remittances or savings), these wage differentials *understate* the consumption gains to movers.

high-productivity country is the over 9,000 deaths in 2024 of people attempting to evade border control while migrating (Missing Migrants Project n. d.).[8]

A harder question concerns the impact of additional movers on the wages of the workers already present. This question is hard, in part because the answer must be "it depends." Because movement is regulated, the number and composition of the foreign-born in any given country is the endogenous aggregation of those who are allowed by the receiving country to be there, either by right or in practice (through lax enforcement). The impact of additional movers on the wages of existing workers must depend on whether the migration constraint was relaxed in a way that produced more movers who were substitutes or complements for existing workers.

Clemens and Lewis (2024) use the fact that US firms' access to foreign-born workers via the H2-B visa for temporary workers is quota-constrained, and hence visas are allocated across firms by lottery. They use this random allocation to estimate the impact of foreign workers on native-born employment and find zero employment effects on the native-born in general and actually positive employment effects in rural areas. This is not a general finding about the employment impact of "migration" because the H2-B visa is constructed to allow domestic firms to hire workers only when native-born workers are not available. Thus, one should expect the expansion at the margin of this program would allow increased employment of workers who are complements for native-born workers by design. Similarly, from 1942 to 1964 the Bracero program in the United States admitted migrants on a temporary basis, mainly to do farm labor. The end of the Bracero program, which reduced the availability of foreign-born agricultural workers in some states and not others, was estimated to have had zero impact on either employment or wages of workers in agriculture in the affected states (Clemens, Lewis, and Postel 2018). Again, the Bracero program was designed to attract workers who were complements in production to US inputs of capital, technology, and native-born labor, and hence the expectation should have been that its elimination would not increase domestic employment.

For the United States, the many studies of the impact of immigration on the wages of native-born have actually created a striking consensus that the effect is really small (NAS 2017). While alternative estimates are slightly on either side of zero (and particularly for subgroups like those with low education), the magnitudes, whether positive or negative, are small.

Instruments to Targets: "Migration" versus "Mobility"

The idea that high-income countries could use movement of workers from other countries to augment their labor force in response to aging seems a political

[8]One field lab experiment assessed the reported willingness of youth in rural West Africa to attempt migration to Europe where the "intervention" was providing accurate information about the probability of dying in route. When provided the "correct" information that the probability of dying was 1 in 4, the self-reported willingness of youths to take the risk *increased* (Bah and Batista 2020).

nonstarter at present. Political pushback against immigration (as it is currently construed) has been strong not only in the United States, but in Germany, France, the United Kingdom, Italy, the Netherlands, and many other countries.

But "immigration" is not the only form of movement of people across borders. When considering such movement, the fundamental policy question is: "Under what terms and conditions (for example, duration of stay, purposes, allowed activities, fiscal entitlements) should our country allow citizens of other countries to be legally present?" Most countries actively attract tourists. Many countries aggressively recruit international students. A more specific policy question is: "Under what terms and conditions are citizens of other countries allowed to be legally present and work (and in what sectors or occupations or even places)?"

There are three distinct types of movers allowed to work.

One type of movers, which is often what is meant by an "immigrant," is people on a pathway to citizenship. This path may be long and contingent, but host societies allow these movers to live and work in a country on the expectation that they form a part of the "future of us."

A second type is "movers of distress," people who the host society allows to move and work out of primarily humanitarian motives. The "mover of distress" is broader than the legalistic concepts "refugee" or "asylum seeker," as these terms tend to be pinned down by specific national or international agreements. While these movers come under different legal provisions, in many cases they are expected to remain in the receiving country for the long term.

A third type of movers is people who move to work on legal authorization that is explicitly time-limited and for which the terms and conditions are narrow, such as limited to a specific occupation. For these movers there is no expectation on either side that this move will be long term or "permanent," or that the mover is on a direct path to citizenship (though it may be a stepping stone). The United States has a number of visa categories of this type: for skilled workers (H-1B), for workers in agriculture (H-2A) or nonagriculture (H-2B), for work as part of "cultural exchange" (such as *au pairs* on a J-1 visa), and for students to work in the United States for a limited period after a master's degree (OPT, an extension of F-1). This type of mover is not an "immigrant" in the usual sense.

I propose an analogy. In the theory of the firm, one can think of the essence of ownership of an asset as residual control rights. In a sophisticated market economy, the owners of an asset can typically enter into a wide variety of contracts for use of their asset without relinquishing any of the residual control rights. As a counterexample, imagine that the only access to use rights of a building was ownership, and that there was no market for leases or rentals that gave access and limited control of a property on a time-limited contractual basis. This would be a massive constraint on the ability of an economy to be productive, because every new enterprise that needed space would need to buy. One source of the high productivity of the rich industrial countries is the ideas of "open access" orders and "market supporting institutions" that create a wide range of contractual forms that allow people (and people as collectives in firms and organizations) to cooperate in the

production of value. I am finishing the revisions to this paper sitting in a room rented by the hour by a coworking space firm that intermediates long leases on office space into a flexible variety of short-term contracts.

The citizens of highly productive democracies have collective control rights over access to their territory that are jealously guarded. Citizens can be viewed as the collective owners of a highly productive space with spatially associated rules, norms, institutions, and fixed capital stocks (infrastructure, buildings, and so on) within which people and firms are able to cooperate in the production of value. The idea of "immigration" as "permanent" movement is that untrammeled labor market access (as a subset of being able to enter into any economic contract) and political residual control rights are deeply tied within the idea of "citizenship." The idea of "labor mobility" is that citizens choose to allow citizens (and collectives, firms, or organizations) to contract with noncitizens to reside and work, but without any implication that conditional access gives the temporary worker a path to the political control rights of citizenship. As labor forces contract, the ability to create value in high-productivity places will be increasingly constrained by legal restrictions that prevent combining core-skill workers into highly productive value chains when physical presence is needed. As a crude analogy, citizens can "rent" access to their highly productive spaces with specified terms and conditions (use rights), but without political control rights to people who would massively gain from access to the wage "place premium" of their country.

The Policy Opportunity

The first column of Table 2 shows the total labor force gap in the UN zero migration scenario, with the assumptions that (1) labor force participation ratios by age and sex are constant at their 2020 levels and (2) countries maintain their 2030 ratio of labor force to over-65 population.

The combination of the wage gains for movers and the numbers of workers needed due to shrinking labor forces relative to the elderly in Table 2 imply massive potential economic gain. Even for low-formal-education or low-skill workers, the typical estimated wage *gain* of a move from a low- to high-productivity country is around $25,000 (in 2017 purchasing power parity dollar) (Pritchett and Hani 2020). If there were 351 million additional movers with an average gain of $25,000, this would increase global wages by $8.7 trillion. To put it another way, keeping the ratio of labor force to over 65 ratios constant would produce wage gains—allowing these magnitudes of mobility would add to wages almost twice the total current GDP of France.

But a "policy opportunity" needs to be a political possibility. However pro-immigration, in the sense of pathways to permanence, you personally might be (and I personally am quite pro-immigration), the labor force gaps are far too large for the combined "permanent" modes of movement ("pathway to citizenship" and "movers of distress") to fully address the gap. If permanent movements alone were allowed to fill the labor force gaps, the magnitudes of movement by 2050 would imply that existing citizens would lose their control of their "future of us."

Table 2
Implications of Meeting Labor Force Gaps with Permanent Migration

	Labor force gap 2050 (millions)	Additional foreign-born in 2050 population if all labor force gap is filled by permanent modes of movement (millions)	Percent of additional permanent migrants (over and above 2020 foreign born) in 2050 population
Column:	I (sorted within region)	II (= col I × 2.1)	III
Spain	23.6	49.5	53.8%
Germany	22.8	47.8	39.5%
United Kingdom	18.1	38.0	36.5%
Italy	17.1	35.9	41.8%
Poland	13.6	28.6	45.9%
France	13.0	27.4	29.8%
Netherlands	5.1	10.6	38.3%
Switzerland	4.5	9.5	53.1%
Austria	3.7	7.8	48.8%
Eight mid-sized	22.2	46.7	41.0%
Nine smallest	6.9	14.4	37.4%
Europe	150.6	316.2	41.0%
United States	107.1	224.9	39.9%
Canada	15.7	32.9	47.0%
North America	122.8	257.8	40.7%
S. Korea	41.3	86.7	66.4%
Japan	23.1	48.6	32.6%
Australia	11.2	23.6	46.2%
New Zealand	2.1	4.4	45.6%
Asia and Pacific	77.8	163.3	47.9%
Total (32 countries)	351.1	737.2	42.2%

Source: Pritchett (2024), Table 3.
Note: The calculations in the first column use the UN Zero Migration scenario, 2020 labor force participation rates, and each country keeping its ratio of labor force to 65 plus constant at its own 2020 value.

Because permanent migrants generally are allowed to move with families, I assume 1.1 additional residents for every migrant working, and hence the estimate in column 2 of total movers needed if all are permanent is just column 1 (needed additional workers) times 2.1. Spain, with a labor force gap in 2050 of 23.6 million people, would need (over and above their existing migrant stock in 2020) 49.6 million migrants by 2050. The United States, for example, would need 224 million additional movers.

In Table 2, column 3 shows the incremental movers from 2020 to 2050 in the total population in 2050. Each of the advanced country regions would have more than 40 percent of their population as additional movers since 2020. Adding this increment to the existing foreign-born in 2020 implies that in most countries the foreign-born would be over 50 percent of the population (Pritchett 2024).

The numbers in Table 2 are not intended to be a "projection" or even a "scenario"—just arithmetic. Of course, one can tinker with the arithmetic by

changing some parameters. Is maintaining the 2020 ratio of labor force to elderly population the appropriate goal, or could countries perhaps make lower support ratios to the elderly work fiscally through cuts in benefits and increases in taxes? Perhaps policy could increase the labor force participation rates? However one tinkers, the stark arithmetic of demographic consequences of low fertility and hence aging poses a question: "Is it likely that voters in the advanced, democratic, rich, and industrialized countries will choose via their democratic processes to allow permanent movers (pathway to citizenship and movers of distress) from now until 2050 such that the ratio of immigrants to foreign-born increases to meet the labor force gaps?" I think the answer is: "No. Nothing even remotely like that is going to happen."

The 2020 percent of foreign-born from countries other than the 32 advanced, democratic, and rich countries (which nets out within-EU movement and other rich-to-rich migration) is under 10 percent in every European country (Pritchett 2024). Yet tensions over immigration are already remaking the politics of Europe. In 2025 anti-migration parties were part of the government in Hungary, Italy, Poland, and the Netherlands (until March 2025). In August 2025, the parties leading the polls in the three largest countries of Europe, Germany, France, and the United Kingdom, were all strongly anti-migration. The idea that politics would change to allow the existing permanent modes of movement to sustain ratios of foreign born in the population to be three to four times higher than their current levels within the next 25 years seems implausible, at best.

Can Rotational Migration Add to the Total Stock of Movers?

There is huge scope for rotational or temporary labor mobility, in which the labor-abundant countries benefit from sending workers on an explicitly time-limited basis to help labor-scarce countries address their labor force gaps. This is not "instead of" immigration, because nearly all countries want to maintain permanent migration channels, using permanent migration to build a bright "future of us" and engaging the "global war for talent" (Kapur and McHale 2005). But there are two fundamental reasons why legal pathways for time-limited movement can expand even when politics are leery or hostile to permanent movement.

First, temporary movement of people to work comes without the same social and political implications of permanent pathways. If resistance to immigration is driven not primarily by economic interests (wages and fiscal implications) but rather by its social and political consequences, and "mass" migration is seen to cause an inevitable and permanent loss of the control by the existing citizens over the "future of us," then allowing individuals to arrive on an explicit (and enforced) time-limited basis is a parallel and additional pathway for movers. In line with the analogy of the owning/renting of an asset above, the possibility of "rental" vastly expands the options for owners, as the rental decision comes with enormously lower and obviously less long-lasting consequences than selling.

Second, the authorization of people to be present to do work can come with terms and conditions about occupations, region of residence, fiscal entitlements, and other matters that would be impossible (and unfair) if imposed on permanent

movers. As labor force scarcities in the high-income countries increase, it will become clear that there are many needed core-skill jobs that cannot attract sufficient citizens. Nearly all countries already have visas that allow for temporary (often seasonal) workers in specific sectors.

When the terms and conditions for temporary presences for work are decided by the host country, then the objections to allowing low-skill workers due to their wage, employment, or fiscal consequences becomes a matter of policy design rather than a necessary consequence of more permanent movers. For instance, it is true, almost by definition, that in a country with a fiscal (tax and benefits) system that is designed to be redistributive, admitting lower-skill foreign workers will have a higher total fiscal cost than admitting high-skill workers. A recent study in the Netherlands by van de Beek et al. (2024) of the net lifetime fiscal contribution of immigrants varied widely across regions. In their base case estimates by country, the lifetime net contribution of an immigrant to the Netherlands from Germany and Austria had a personal primary income about 89 percent of a Dutch native's and hence was roughly a fiscal wash (lifetime €23,000). In contrast, immigrants from poor countries, like Pakistan, had only about 44 percent the lifetime earnings and hence had a net lifetime fiscal contribution of negative €238,000. But if workers are only present to work for some months to at most a few years, and arrive without additional family members, then core-skill workers can be a fiscal wash or even, depending on the design of the contract, a positive fiscal contribution.

To be politically acceptable, perhaps even popular, the implementation of policies of temporary mobility will need to be, and be seen to be, orderly and legitimate. Two key issues are that a policy of allowing time-limited workers needs to have effective enforcement of return and high compliance. If voters see routine noncompliance and overstay—and the use of temporary and conditional legal access abused as a pathway to permanence—this obviously discredits the entire premise. Similarly, if voters see the use of foreign workers as means of undermining the conditions of work (wages, safety) of citizens or as a means of exploiting movers (for example, by not delivering on contractual obligations to workers) this delegitimizes the idea of mobility.

Effective implementation of massively scaled labor mobility is going to require public and private cooperation between sending and host countries. The five core functions required to mediate a successful move are: (1) *recruitment*, attracting people in a sending country to the opportunity; (2) *preparation,* so that workers arrive in the host country prepared to do the work for which they have contracted; (3) *placement* mechanisms to match movers to jobs that do not create the conditions of exploitation through a firm-linked presence; (4) *protection,* so that movers have ready access means of preventing abuse and exploitation; and (5) *return,* so that there is compliance with the time-limited residence in the host country. Creating an industry that moves people at this scale people is a huge challenge, but not radically more so than the complex cooperation of the airline industry that allows billions of passengers to fly safely or the complex industry of tourism, whereby 1.4 billion people visit

other countries every year. These advanced industrial countries have high capability governments and successfully regulate a huge range of complex industries (Pritchett 2023). As a final analogy, the US experience with passage and repeal of Prohibition, the attempt to ban the production and sale of all alcoholic beverages, showed that there can be a "bridge too far" such that enforcement will fail but that legalization of transactions citizens want to engage in (like hiring foreign workers) can bring order and control.

Conclusion

The simple arithmetic of demography—with births, deaths, and everyone between those two events aging by exactly one year every year—makes the age structure of population, even three decades ahead, quite predictable. Economic or political changes can happen very fast with unpredicted timing, like the post-1978 rise of China, the collapse of the Soviet Union, or the 2008 global financial crisis. But the demographic implications of fallen fertility for the age structure of the population (without migration) out to at least 2050 are already, absent catastrophe, destiny.

Thus, we know with considerable confidence that in the next few decades, nearly all of the expansion in the labor force aged population will happen in countries with low levels of GDP per worker, while the countries with the lowest ratios of working-aged population to retirees will mostly be in the high-income countries of the world (plus China and Russia). One implication of this largely uncontested demographic arithmetic is that the largest, and growing, opportunity for known and reliable policy-driven economic gains for the global population is the creation of legal pathways for safe, regular, orderly, and large-scale movement of labor force aged people from the low-productivity countries where they are going to be born to the high-productivity (and hence high-wage) countries with shrinking labor forces and expanding numbers of elderly.

It seems politically impossible to meet the labor force needs of the aging economies with immigration based exclusively on a "pathway to citizenship" and "movers of distress." Even if these "permanent" migration pathways are expanded to the limits of political acceptability, there remains considerable scope for a vast expansion of the temporary and rotational pathways for legally authorized movement of people to work.

■ *I would like to thank Michael Clemens for inspiration and impetus, Rebekah Smith and the Labor Mobility Partnerships team for wisdom gleaned from practical engagement, Timothy Taylor for impressive (and compressive) editing, the Editorial Board and Sam Steyer for comments, and Patrick Collison for pointing me to the study of fiscal impacts.*

References

Ashenfelter, Orley, and Stepan Jurajda. 2001. "Cross-Country Comparisons of Wage Rates: The Big Mac Index." Princeton University Economics Department Working Papers 2001–7.

Bah, Tijan L., and Catia Batista. 2020. "Understanding Willingness to Migrate Illegally: Evidence from a Lab in the Field Experiment." NovaAfrica Working Paper 1803.

Barro, Robert and Jong-Wha Lee. 2013. "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics* (104): 184–98. Accessed May 2025.

Bergsvik, Janna, Agnes Fauske, and Rannveig Kaldager Hart. 2021. "Can Policies Stall the Fertility Fall? A Systematic Review of the (Quasi-) Experimental Literature." *Population and Development Review* 47 (4): 913–64.

Büttner, Nicolas, Michael Grimm, Isabel Günther, Kenneth Harttgen, and Stephan Klasen. 2024. "The Fertility Transition in Sub-Saharan Africa: The Role of Structural Change." *Demography* 61 (5): 1585–611.

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

Clemens, Michael A., and Ethan G. Lewis. 2024. "The Effect of Low-Skill Immigration Restrictions on US Firms and Workers: Evidence from a Randomized Lottery." NBER Working Paper 30589.

Clemens, Michael A., Ethan G. Lewis, and Hannah M. Postel. 2018. "Immigration Restrictions as Active Labor Market Policy: Evidence from the Mexican Bracero Exclusion." *American Economic Review* 108 (6): 1468–87.

Clemens, Michael A., Claudio E. Montenegro, and Lant Pritchett. 2019. "The Place Premium: Bounding the Price Equivalent of Migration Barriers." *Review of Economics and Statistics* 101 (2): 201–13.

Dupas, Pascaline, Seema Jayachandran, Adriana Lleras-Muney, and Pauline Rossi. 2025. "The Negligible Effect of Free Contraception on Fertility: Experimental Evidence from Burkina Faso." *American Economic Review* 115 (8): 2659–88.

Eberstadt, Nicolas. 2022. "The De-Population Bomb." Interview with Peter Robinson, Uncommon Knowledge podcast series, September 12. https://www.youtube.com/watch?v=uNdnlrkx-wg.

Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer. 2015. "The Next Generation of the Penn World Table" *American Economic Review* 105 (10): 3150–82. Available for download at www.ggdc.net/pwt.

GBD (Global Burden of Disease 2021 Fertility and Forecasting Collaborators). 2024. "Global Fertility in 204 Countries and Territories, 1950–2021, with Forecasts to 2100: A Comprehensive Demographic Analysis for the Global Burden of Disease Study 2021." *The Lancet* 403 (10440): 2057–99.

Goldin, Claudia. 2025. "Babies and the Macroeconomy." *Economica* 92 (367): 675–700.

Goldstone, Jack. Forthcoming. *10 Billion: How Population Will Change the World in the 21st Century*. Oxford University Press.

Goldstone, Jack A. 2010. "The New Population Bomb: The Four Megatrends That Will Change the World." *Foreign Affairs* 89 (1): 31–43.

Günther, Isabel, and Kenneth Harttgen. 2016. "Desired Fertility and Number of Children Born across Time and Space." *Demography* 53 (1): 55–83.

ILOSTAT. 2024. *Labour force participation rate*. International Labour Organization. https://ilostat.ilo.org/data/snapshots/labour-force-participation-rate/.

Jones, Charles I. 2022. "The End of Economic Growth? Unintended Consequences of a Declining Population." *American Economic Review* 112 (11): 3489–3527.

Kapur, Devesh, and John McHale. 2005. *Give Us Your Best and Brightest: The Global Hunt for Talent and Its Impact on the Developing World*. Center for Global Development.

K. C., Samir, Moradhvaj Dhakad, Michaela Potančoková, Saroja Adhikari, Dilek Yildiz, Marija Mamolo, Tomas Sobotka, et al. 2024. "Updating the Shared Socioeconomic Pathways (SSPs) Global Population and Human Capital Projections." International Institute for Applied Systems Analysis Working Paper WP-24-003.

Kohler, Hans-Peter, Francesco C. Billari, and José Antonio Ortega. 2002. "The Emergence of Lowest-Low Fertility in Europe during the 1990s." *Population and Development Review* 28 (4): 641–80.

McKenzie, David, Steven Stillman, and John Gibson. 2010. "How Important Is Selection? Experimental Vs. Non-Experimental Measures of the Income Gains from Migration." *Journal of the European Economic Association* 8 (4): 913–45.

McNamara, Robert S. 1977. "Address to the Massachusetts Institute of Technology by Robert S. McNamara President, World Bank (English)." Presidential speech, World Bank Group. http://documents.worldbank.org/curated/en/701961478757089433.

Missing Migrants Project. n. d. "Deaths During Migration Recorded since 2014, by Region of Incident." International Organization for Migration. https://missingmigrants.iom.int/data (accessed November 2025).

Mobarak, Ahmed Mushfiq, Iffath Sharif, and Maheshwor Shrestha. 2023. "Returns to International Migration: Evidence from a Bangladesh-Malaysia Visa Lottery." *American Economic Journal: Applied Economics* 15 (4): 353–88.

NAS (National Academies of Sciences, Engineering, and Medicine). 2017. *The Economic and Fiscal Consequences of Immigration*. The National Academies Press.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business and Economic Statistics* 37 (2): 187–204.

Pritchett, Lant H. 1994. "Desired Fertility and the Impact of Population Policies." Population and Development Review 20 (1): 1–55. https://doi.org/10.2307/2137629.

Pritchett, Lant. 2018. "Alleviating Global Poverty: Labor Mobility, Direct Assistance, and Economic Growth." HKS Working Paper RWP18-013.

Pritchett, Lant. 2023. "The Political Acceptability of Time-Limited Labor Mobility: Five Levers Opening the Overton Window." *Public Affairs Quarterly* 37 (3): 284–306.

Pritchett, Lant. 2024. "Rotational labor mobility is the biggest global economic opportunity" Unpublished. https://lantpritchett.org/wp-content/uploads/2024/07/potential-for-labor-mobility-biggest_firstcomplete.pdf.

Pritchett, Lant. 2026. *Data and Code for: "Global Labor Mobility between Shrinking and Growing Labor Forces."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E239755V1.

Pritchett, Lant, and Farah Hani. 2020. *Oxford Research Encyclopedias*. Oxford University Press.

Pritchett, Lant, and Martina Viarengo. 2012. "Why Demographic Suicide? The Puzzles of European Fertility." *Population and Development Review* 38: 55–71.

Senderowicz, Leigh, and Nicole Maloney. 2022. "Supply-Side versus Demand-Side Unmet Need: Implications for Family Planning Programs." *Population and Development Review* 48 (3): 689–722.

Spears, Dean, and Michael Geruso. 2025. *After the Spike: Population, Progress and the Case for People*. Simon and Schuster.

The Economist. 2024. "Why paying women to have more babies won't work." March 23.

UN (United Nations). 2024. *World Population Prospects*. Department of Economic and Social Affairs, Population Division.

van de Beek, Jan, Joop Hartog, Gerrit Kreffer, and Hans Roodenburg. 2024. "The Long-Term Fiscal Impact of Immigrants in the Netherlands, Differentiated by Motive, Source Region and Generation." IZA Discussion Paper 17569.

Weil, David N. 2024. "Replacement Fertility Is Neither Natural nor Optimal nor Likely." Unpublished.

Journal of Economic Perspectives—Volume 40, Number 1—Winter 2026—Pages 93–114

Labor Market Power: From Micro Evidence to Macro Consequences

David Berger, Kyle Herkenhoff, and Simon Mongey

In a textbook model of perfect competition in labor markets, workers' wage compensation is determined by the interaction of demand from atomistic firms and supply of individual workers. But in real-world labor markets, employment and compensation are often determined in a setting where the demand-side of a local labor market for a certain industry is dominated by a few large local employers—or sometimes, even a single large employer. According to the US Bureau of Labor Statistics (2025), out of 132 million private industry workers in March 2025, 16 million worked at establishments with 1,000 or more workers, another 8 million at establishments with 500–999 workers, and 12 million more at establishments with 250–499 workers. (Remember, an establishment is a physical location where a business operates and hires; a large firm will consist of multiple establishments.) A substantial and growing body of recent research has examined the microeconomic and macroeconomic implications of the way these large firms exercise their labor market power.

This essay argues that understanding labor market power requires a macro perspective. Well-identified micro estimates are indispensable for documenting wage-setting power and pinning down key elasticities. But labor market power is concentrated among large employers in local markets that are themselves highly concentrated, and in these settings micro evidence alone is not sufficient. Once

■ *David Berger is Professor of Economics. Duke University, Durham, North Carolina. Kyle Herkenhoff is Associate Professor of Economic, University of Minnesota, Minneapolis, Minnesota. Simon Mongey is Monetary Advisor, Federal Reserve Bank of Minneapolis, Minneapolis, Minnesota. Their email addresses are david.berger@duke.edu, kyle.herkenhoff@gmail.com, and simon.mongey@frb.mpls.org. The views in this paper do not represent those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251456.

dominant employers are present, three mechanisms amplify the aggregate consequences of monopsony. *Granularity* (or the pure largeness of the firm) means that their wage and hiring decisions mechanically move market-wide averages because these firms account for a large share of employment. *Strategic interaction* arises because dominant employers internalize their outsized footprint and understand that their choices directly shift wages and employment in the broader market. *Misallocation* follows when the largest and most productive employers restrict hiring, pushing workers toward less efficient firms and thereby magnifying welfare losses.

These macro channels pose significant challenges to existing theoretical and empirical "micro approaches" that rely on small, atomistic firm behavior. Yet, the markets in which the assumptions underpinning the micro approach fail are precisely the markets that motivate the most pressing policy debates—noncompetes, no-poach agreements, and merger enforcement—and in which scope for monopsony is most severe. To make progress, we propose a macro approach in which researchers build and estimate theoretical models of granular firms harnessing both aggregate data and micro empirical estimates. We argue that the macro channels invert conventional policy prescriptions. The most effective interventions directly target dominant employers—through antitrust enforcement, limits on restrictive contracting such as noncompete clauses, and related tools—while broad-based policies like minimum wages are often poorly aimed at the firms exercising the most market power.

We begin by defining concentration in labor markets using the Herfindahl-Hirschman Index (HHI) and documenting the extent to which workers face concentrated employer demand. We then examine evidence that in concentrated labor markets, workers are paid significantly less than their marginal product—across different approaches and data sources, the median firm pays its workers between 70 and 80 percent of their marginal product. We discuss how the three macro mechanisms propagate these wage markdowns into aggregate welfare losses: across a range of methodologies, these losses total more than 7.6 percent of a typical household's lifetime consumption, and wage gains from eliminating labor market power can exceed 30 percent in the most concentrated markets (Berger, Herkenhoff, and Mongey 2022a; Jarosch, Nimczik, and Sorkin 2024). We then discuss the conceptual limitations of the traditional theoretical and empirical micro approach, which requires an assumption of small, atomistic firms. We outline the macro approach to monopsony, explaining how to harness both aggregate and identified micro elasticities to learn about spillovers and granular firm behavior. Finally, we examine the policy implications of this macro perspective, emphasizing how concentration among large employers changes which interventions are most effective.

Measuring Labor Market Concentration

To define a labor market, we start with the idea that these markets happen within the geographic area in which people live and work (Berger, Herkenhoff, Mongey 2022a). We use the idea of a "commuting zone" developed by researchers at

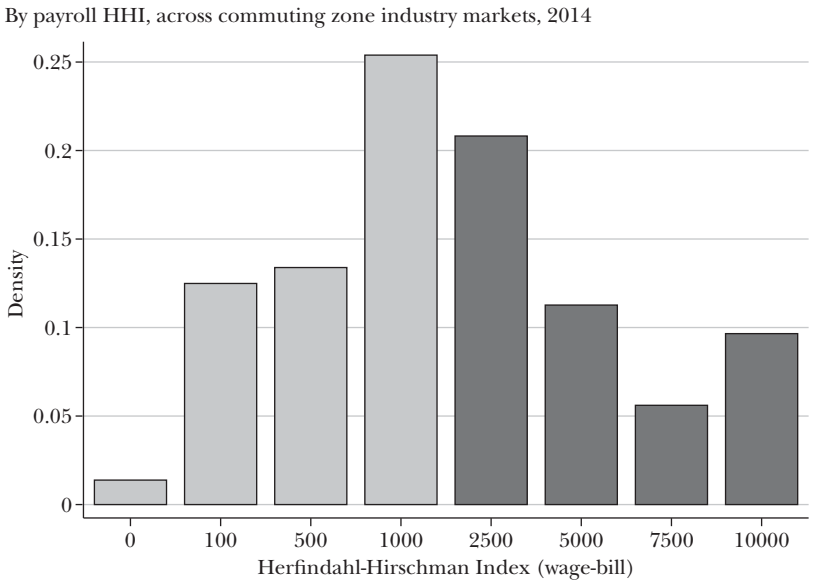
the US Department of Agriculture (USDA Economic Research Service 2025), who use county-level data to divide the US economy into 592 commuting zones, which as the name suggests are determined by the extent of actual commuting patterns within these areas. In addition, most workers have a degree of industry-specific knowledge and thus tend to look for jobs within a certain industrial sector. Here, we focus on 96 industries as classified at the three-digit level in the North American Industry Classification System (NAICS), which is meant to strike a balance between the more general classification of 20 sectors at two-digit level and the 1,012 sectors at the highly detailed six-digit level.

To measure concentration across labor markets, we adopt the well-known Herfindahl-Hirschman Index. In the labor market context, the HHI is computed by determining the market share of each company in a labor market—that is, their payroll divided by the total payroll of their competitors in that three-digit industry sector in a given commuting zone—then squaring and summing those shares. In several theories of noncompetitive labor markets, HHIs emerge as a sufficient statistic for a market's competitiveness. For example, Berger, Herkenhoff, and Mongey (2022a) show that because wage-setting power in oligopsony models is a direct function of a firm's share of total payroll, the correct statistic is a wage-bill-weighted HHI. At an intuitive level, when a few firms dominate employment and hiring in a certain commuting zone and industry labor market, job-seeking workers repeatedly encounter these same employers, and what matters is how much of total wages they are paying. Jarosch, Nimczik, and Sorkin (2024) provide a complementary micro foundation for the HHI, showing that the HHI is a natural summary of labor market competitiveness when workers search for jobs in concentrated market.

Figure 1 shows that many US labor markets, as defined by commuting zone and three-digit industry sector, are quite concentrated. Each bar shows the share of total labor income in the economy paid in markets with particular concentration levels. For example, the tallest bar shows that 25 percent of total labor income is paid in local labor markets with an Herfindahl-Hirschman Index of around 1000, which corresponds to the level of competition generated by ten equally-sized firms. While the average labor market has more than 100 firms, employment and wage payments are highly concentrated among dominant firms. The US antitrust authorities at the Federal Trade Commission and the Antitrust Division of the US Department of Justice often use an HHI of 2500 as a rough rule-of-thumb for whether they should be concerned about the effects of a proposed merger. As the dark gray bars show, approximately 40 percent of total wage payments in the US economy are made in labor markets that are at or above this threshold.

Similar measures of concentration emerge with other measurement approaches. Azar, Marinescu, and Steinbaum (2022) compute Herfindahl-Hirschman Indexes using job postings rather than wage payments, defining market share as the share of vacancies or applications received in local markets. Despite these methodological differences, they find approximately 40 percent of markets are highly concentrated, in the sense that they exceeding the commonly used HHI threshold of 2500. A high degree of labor market concentration appears to be a robust empirical phenomenon.

Figure 1
Labor Market Concentration and Wage Markdowns



Source: Berger, Herkenhoff, and Mongey (2022a).

Labor Market Power at the Firm Level

Traditional Sources of Labor Market Power

Real-world labor markets are far from a textbook example of supply and demand for a homogeneous product being traded in a frictionless and full-information market. As a result, there are several reasons why employers have a degree of market power even without raising issues related to large or concentrated employers.

First, jobs differ in amenities, tasks, schedules, and workplace conditions, and workers place heterogeneous value on these characteristics. As a result, no two firms are perfect substitutes from the workers' perspective, giving even atomistic firms a degree of wage-setting power. Because workers sort toward employers whose characteristics they value, each firm faces its own upward-sloping labor supply curve, rather than competing in a common spot market. Several issues are interrelated here: worker sorting, compensating differentials, rent sharing, and imperfect competition, but the key point is that differentiation across jobs gives firms market power. Recent papers support the quantitative importance of this mechanism. Lamadon, Mogstad, and Setzler (2022) estimate a model using matched US worker-employer tax data from 2001 to 2015 that allows workers to differ in their preferences over employers and firms to differ in productivity and hiring behavior. They show that these forms of heterogeneity are not enough to replicate the observed wage and

mobility patterns unless firms also have market power in setting wages. Similarly, Card, Heining, and Kline (2013) document large firm-specific wage components in West Germany, underscoring the central role of employers in wage determination.

Second, a job search requires time and effort, and mobility to new jobs is often costly (buying or selling a home, changing children's schools, and other costs), so workers cannot move seamlessly between jobs—even when those jobs are otherwise functionally identical. The canonical search model of Burdett and Mortensen (1998) demonstrates how these frictions naturally generate labor market power. Firms recognize that workers cannot instantly switch employers, creating room for monopsonistic wage-setting below competitive levels. Some firms choose to pay higher wages to attract workers more quickly, while other firms opt for lower wages, accepting higher turnover and slower hiring. Search frictions effectively create bilateral monopoly situations and generate the kind of firm-level wage-setting power emphasized in Manning's (2003) synthesis of the modern monopsony literature.

These search frictions provide a micro-foundation for why the labor supply curve facing each is upward sloping, even when firms are small. Our focus in this essay is on how these traditional sources of monopsony interact with granularity and market structure when some firms are very large. Recent work such as Jarosch, Nimczik, and Sorkin (2024) explicitly combines frictional search with big firms in concentrated labor markets, showing how granular employers use their position in the job ladder to shape wages. Our "macro approach" can be viewed as building on this search tradition—using rich worker-firm data and models with large employers to measure the resulting wedges and trace out their aggregate consequences.

Third, workers will have incomplete information about their "outside options"—that is, the alternative jobs that are available to them in the labor market and how much those jobs would pay. For a review of early theoretic foundations for monopsony power that stem from incomplete information about job offers, a useful starting point is Mortensen and Pissarides (1999). Recent incarnations appeal to the theory of "rational inattention" to motivate monopsony power—that is, a typical worker will not find it worthwhile to spend the ongoing time and energy needed to gather information on evolving "outside options" (Cheremukhin, Restrepo-Echavarria, and Tutino 2020; Cheremukhin and Restrepo-Echavarria 2025). In a survey of German workers, Jäger et al. (2024) find that most workers wrongly believe that alternative wages are very close to present wages. For example, "workers that would experience a 10% wage change if switching to their outside option only expect a 1% change." Such biased and limited information about outside options exacerbates labor market power for employers.

Quantitative Measures of Labor Market Power

In product markets, a firm with market power may choose to reduce its output, increase prices, and reap higher profits. Similarly, a firm with labor market power may strategically decide to reduce its labor demand in order to pay lower wages. Recent quantitative theories of labor market power focus on how large firms in concentrated

markets can exploit labor market imperfections to extract worker rents through monopsony power. For example, Berger, Herkenhoff, and Mongey (2022a) study oligopsonistic competition and show that larger, more granular firms as measured by market share will markdown wages much more than smaller atomistic firms.

However, measuring these markdowns—that is, the gap between wages and the marginal revenue product of workers—is not straightforward. Marginal revenue is almost never directly observable; instead, it must be estimated from production functions.

In one prominent study of markdowns, Yeh, Macaluso, and Hershbein (2022) use data on US manufacturing firms from the Census of Manufactures (comprehensive coverage every five years) and the Annual Survey of Manufactures (partial coverage, not including all small firms, but available annually). With data on revenue and inputs of labor, materials, capital, and energy, they estimate plant-level production functions from 1976 to 2014, which in turn allows them to calculate the marginal revenue product of labor. They find that workers are paid about 65–73 percent of their marginal output. This markdown varies substantially across firms. As Figure 2 shows in panel A, when the plants with larger shares of local employment (measured at the county by three-digit industry level), the markdown is systematically greater. In another study of markdowns, Chan et al. (2023) use Danish data and a model to estimate firm-level distributions of productivity, worker ability, markdowns, pass-through, and labor-supply elasticities. They find that the median worker receives about 79 percent of their marginal product in pay as shown in Figure 2, panel B.

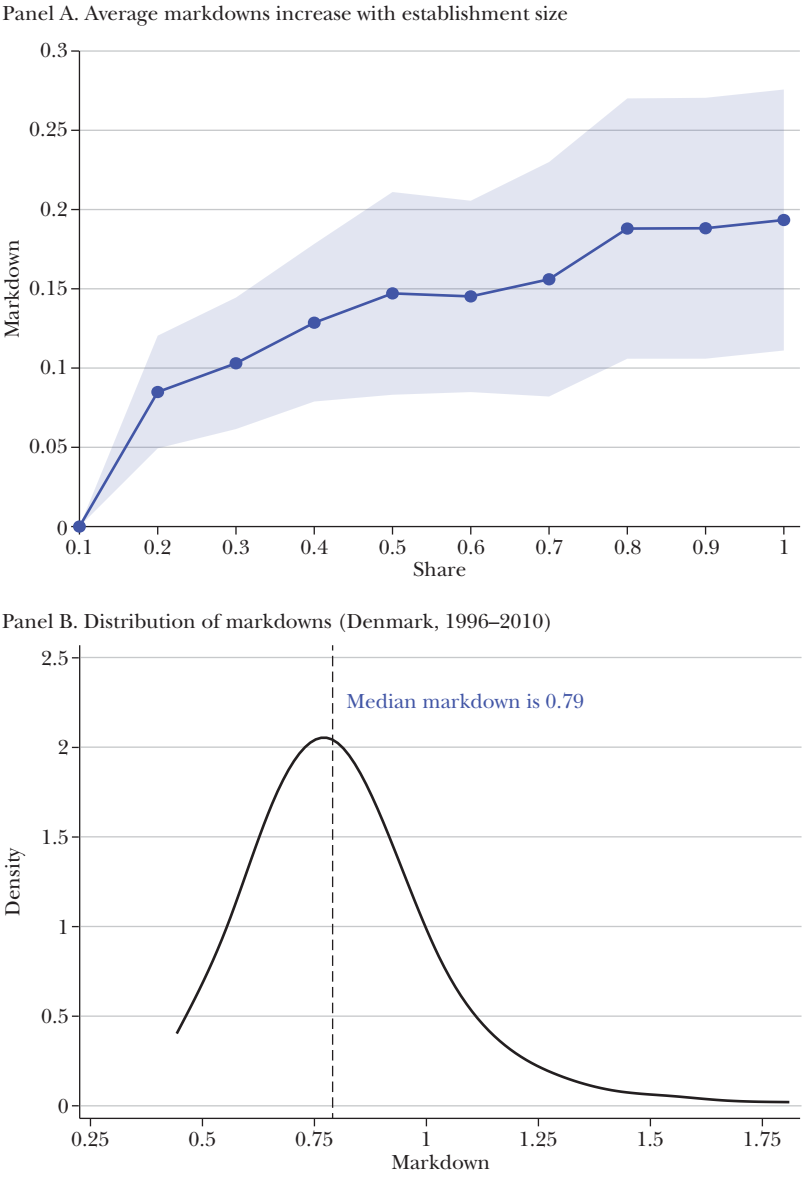
When workers have limited outside options, firms gain bargaining power. In frictional labor markets, this power is amplified for large employers who control a substantial share of available jobs. Jarosch, Nimczik, and Sorkin (2024) formalize this mechanism using Austrian data: larger firms that dominate local job listings can credibly threaten to exclude workers from future job opportunities if they reject the firm's terms. This gives granular employers leverage to markdown wages more aggressively than atomistic firms.

This essay is not a literature review. But across these studies—and the broader literature they cite—a consistent pattern emerges: workers in concentrated labor markets typically receive 70–80 percent of their marginal product, with the largest markdowns occurring at the biggest firms.

When Labor Market Power Cascades to the Macroeconomy

When large employers dominate labor markets, their decisions do not just affect markdowns paid to their own workers. Their actions may cascade through entire economies. In this section, we argue that dominant firms' actions propagate through the economy via three mechanisms: granularity, strategic actions, and misallocation. For each mechanism, we discuss its theoretical basis and supporting empirical evidence.

Figure 2
Evidence on Wage Markdowns



Source: Panel A is from Yeh, Macaluso, and Hershbein (2022). Panel B is from Chan et al. (2023).

Granularity

In common usage, "granular" refers to something small or divided into many parts. However, in the research on labor market power, "granular" is in contrast to the idea of truly atomistic markets—in other words, "granular" means big enough

for individual actions to matter. This difference is not just a theoretical curiosity: as Figure 1 shows, in 11 percent of US labor markets, a single employer accounts for at least 11 percent of the wage bill, and the average labor market has concentration equivalent to just nine equally-sized firms. When firms are this large, their actions—or shocks and policies that affect them—can influence market-, state-, and even national-level aggregates.

What does it mean to be a granular firm in a labor market? Formally, if firm i commands wage-bill share s_i, then a 1 percent wage change at that granular firm moves the average market wage by s_i percent, regardless of any strategic considerations. Consider an average labor market in which the Herfindahl-Hirschman Index corresponds to nine equally-sized firms, each hypothetically controlling 11 percent of the market. If, say, Amazon employs 11 percent of the "Warehousing and Storage" workers in a commuting zone (a three-digit industrial code), its wage decisions mechanically move the market average wage by 11 percent.

The pure granularity effect has a long and well-explored literature in product markets. For example, Gabaix (2011) argued that idiosyncratic shocks to the 100 largest US firms explain one-third of variation in output growth. The same logic applies to labor markets. Kovalenko, Schnabel, and Stüber (2022) adapted the approach of Gabaix (2011) to show that the largest firms in Germany are the main drivers of aggregate employment dynamics. Esquierro and da Silva (2024) provide a similar analysis for Brazil, finding that public firms are important in understanding granularity in that context.

Strategic Interaction

In an extreme example with identical workers and a single firm, the monopsony firm will set wages to the workers' common outside option and extract all rents from workers. Early theoretical models of oligopsony labor markets (Bhaskar, Manning, and To 2002; Boal and Ransom 1997) showed how to apply the intuition of this extreme example to markets with a handful of nonatomistic firms that compete strategically for labor. More recent research brings this core oligopsony idea to the data by creating quantitative models that capture key features of high-quality microdata, including rich heterogeneity across firms, workers, and labor markets. When a granular firm competes against other granular firms, these strategic actions are anticipated and responded to. Through this process of market level interaction, the labor market inherently begins to become coordinated around these actions.

A large firm—one that accounts for a meaningful share of local job opportunities—can shape outcomes through several strategic channels. Because its hiring moves the market as a whole, it may choose to restrain its employment in order to keep wage pressure down. This strategic use of hiring to influence market-wide conditions is a core mechanism emphasized in oligopsonistic competition models such as Berger, Herkenhoff, and Mongey (2022a). Understanding that workers' outside options are limited, a granular firm may bargain more aggressively to reduce pay. Jarosch, Nimczik, and Sorkin (2024) show that in frictional labor markets, larger firms controlling a greater share of job listings can credibly

threaten to exclude workers from future job opportunities if they reject the firm's wage offers. Finally, because their actions affect policy and other economic actors, granular firms can take indirect actions to suppress wages, including confronting unions (Azkarate-Askasua and Zerecero 2025) or engaging in tacit or explicit collusion (Brooks et al. 2019; Gottfries and Jarosch 2025b).

Whether firms take strategic actions in labor markets is hotly debated. A first approach points to patterns in markdowns as evidence of strategic interactions. Using a structural model and empirical tools, based on comparing the employment and wage responses of firms that have experienced changes in state-level corporate tax laws to those that have not, Berger, Herkenhoff, and Mongey (2022a) find that larger firms within local labor markets have wider markdowns. Our structural model interprets this as evidence of strategic actions. If large firms were not taking strategic actions to exploit their granular role in the labor market, then they would have the same markdown as small firms. On the other hand, a study of a job board for high-paid tech workers rejects strategic oligopsonistic competition in favor of straightforward monopsonistic competition (Roussille and Scuderi 2023). This finding may be surprising, given that high-paid tech workers are a segment of the labor market that has been explicitly subject to anticompetitive behavior between its largest firms via collusive nonpoaching agreements (Posner and Zheng 2025). The tension between these results demands further advances in measurement.

A second approach to testing for strategic actions is to examine spillovers in local labor markets. In an influential study, Staiger, Spetz, and Phibbs (2010) show that strategic responses are large following legislated wage increases at a big firm in the labor market for nurses: Veterans' Administration (VA) hospitals. For every 10 percent increase in wages at a VA hospital, nearby non-VA hospitals increase their wages by 1.28 percent. More recently, Derenoncourt et al. (2021) examined strategic responses to Amazon's 2018 introduction of a $15 voluntary minimum wage. They document that large national competitors quickly followed suit, entirely consistent with large firms best-responding to each other. However, the authors find limited evidence of local wage spillovers to smaller competitors. In contrast, Ashenfelter and Jurajda (2025) study the same Amazon wage shock but focus on McDonald's restaurants—workers in entry-level, high-turnover jobs with little firm-specific skill. They find that McDonald's restaurants within 10 kilometers of Amazon fulfillment centers raised wages significantly in response to Amazon's policy, with larger increases in highly concentrated and low-wage local labor markets.

Some of this tension may stem from differences in how labor markets are defined. Strategic interaction only makes sense relative to a clearly defined labor market—the set of employers who actually compete for the same workers. In empirical work, this situation is typically approximated by geographic-industry or geographic-occupation cells, but competitive boundaries vary substantially across settings. Hospitals compete locally; large retailers set wages nationally; tech firms compete for specialized workers across broad geographies. The mixed spillover evidence may reflect this heterogeneity. Understanding when and where strategic interactions matter requires careful attention to market definition and the degree

of substitutability among competing employers. More work is needed to understand when and where strategic interactions drive wage setting, particularly across different types of labor markets and worker skill levels.

A third empirical approach studies events that exogenously increase employer concentration, most notably mergers, on firms' actions in a direct way. While these settings do not directly test strategic wage setting, they offer credible evidence on how firms respond when their labor market power expands. Arnold (2020) analyzes mergers across the entire US economy using Census data and finds that consolidation reduces wages, with larger effects in markets that were already concentrated. Prager and Schmitt (2021) examine hospital mergers and similarly document wage declines for hospital workers following consolidation. These results demonstrate that when concentration rises, firms strategically act to cut wages—behavior inconsistent with a competitive model where employers are atomistic price-takers.

A fourth approach directly analyzes legal agreements, case studies, and rulings to argue that strategic actions are important. Posner (2021) reviews a number of influential cases brought against employers over the last century, while also highlighting a number of ways in which antitrust enforcement has ignored arguably per se illegal strategic behavior on the side of employers. In particular, Starr and Smith (2019) discuss the prevalence of firms that require employees to sign noncompetes and no-poach agreements (a topic Starr explores in greater depth in his paper in this symposium). By 2014 roughly 50 percent of workers earning above $150,000 were bound by noncompetes, while 14 percent of those earning below $40,000 were likewise bound. By 2016, 60 percent of large firms were using no-poach agreements, including the Apple and Google agreements mentioned earlier. Regardless of their legality, the law and economics literature provides complementary evidence on widespread strategic actions by firms in concentrated labor markets.

Misallocation

In an efficient economy, high-productivity firms will attract additional productive resources—like labor, capital, and land—allowing these firms to expand. Relative to this efficient allocation of resources, if large, productive firms take actions to reduce employment and wages in labor markets, efficiency will be reduced as resources are misallocated.

Consider a local economy in which the engineering productivity of large Tech Firm G is twice that of small startups. In a competitive market, Firm G would continue hiring engineers until wages equal their marginal product. But with market power, profit-maximizing Firm G will restrict hiring to keep wages low. The talented engineers that would work at Firm G in an efficient economy instead work for lower wages at less productive startups, where their skills generate half the value. Overall economic productivity is lower as a result.

Large markdowns at the most productive firms are especially distorting. The studies mentioned earlier, including Yeh, Macaluso, and Hershbein (2022) on US manufacturing plants, establish that large firms have larger markdowns. This

result holds true in a number of different countries and industries, suggesting wide scope for reallocation to improve economic efficiency.

Do policies that reduce monopsony power actually induce such reallocation? Dustmann et al. (2022) document strong patterns of reallocation from small to large firms in Germany following their large minimum wage hike in 2008. Workers moved toward more productive employers, suggesting that reallocation may be an important source of gains from minimum wages and other policies that address monopsony. More broadly, Carlsson, Messina, and Skans (2021) demonstrate that policies affecting firm-level shocks can drive substantial reallocation in the Swedish labor market, underscoring that labor market frictions shape how workers sort across firms of varying productivity.

Measuring the Aggregate Effects of Labor Market Power

The evidence reviewed above shows that workers in concentrated labor markets are typically paid well below their marginal product, with the largest markdowns occurring at the biggest firms. But the implications of labor market power extend well beyond direct pay suppression. When large employers dominate local labor markets, their decisions reshape outside options throughout the market and affect how labor is allocated across firms. We previously documented empirical patterns consistent with granularity, strategic interaction, and misallocation. In this section, we examine how incorporating these three mechanisms into quantitative models substantially amplifies the estimated aggregate costs of labor market power compared to models with atomistic firms.

A natural benchmark comes from a model that captures direct wage suppression, but without strategic interaction among large firms. In a study mentioned earlier, Lamadon, Mogstad, and Setzler (2022) estimate such a structural model using matched US worker-firm tax records from 2001 to 2015. In their model, workers differ in their preferences over employers and firms differ in productivity and hiring behavior. This rich heterogeneity generates meaningful wage dispersion and sorting, but firms behave as atomistic wage-takers: even large employers take market wages as given rather than recognizing their ability to influence them. Using this framework, they estimate that eliminating labor market frictions and distortions would raise output by 3 percent and welfare by 5 percent. These losses reflect direct markdowns—firms paying workers less than their marginal product—but not the amplification that arises when dominant firms can move market wages or strategically suppress hiring. Lamadon, Mogstad, and Setzler therefore provide a lower bound on the aggregate costs of monopsony power.

Allowing for granular firms that recognize their influence on market conditions substantially increases the implied macro losses. Berger, Herkenhoff, and Mongey (2022a) develop a general-equilibrium model in which large, granular employers internalize that their wage and hiring decisions affect market wages—and anticipate the responses of other large firms. Using variation from state corporate tax changes

to discipline the model, they find that eliminating distortions raises output by 20.9 percent and welfare by 7.6 percent. Output gains are seven times larger than the Lamadon, Mogstad, and Setzler (2022) benchmark, while welfare gains are 50 percent larger. This gap captures the contribution of strategic interaction: if large firms behaved like atomistic competitors, their markdowns would resemble those of small firms rather than being systematically wider.

A major reason the aggregate losses rise so sharply in the Berger, Herkenhoff, and Mongey (2022a) framework is misallocation. When dominant firms restrict hiring to keep wages low, workers who would be most productive at those firms end up at less productive employers. The empirical patterns documented earlier—larger markdowns at the largest and most productive firms—mirror this mechanism. In this paper, we show that misallocation accounts for about 60 percent of the output gap between the efficient allocation and the observed allocation with monopsony power. In other words, much of the aggregate cost arises not from the lower wages themselves but from the fact that workers end up in the wrong jobs.

Labor market power also shapes the wage distribution. In a model in which granular firms in a concentrated local market do not compete against themselves to offer higher wages, Jarosch, Nimczik, and Sorkin (2024) find based on Austrian data that eliminating labor market power would raise wages by 2.6 percent on average, and 30 percent in the most concentrated markets. High earners gain the most from eliminating monopsony, because they are disproportionately employed at the largest firms with the widest markdowns. These distributional effects run counter to the view that monopsony primarily hurts the lowest paid, highlighting significant policy implications that differ markedly from standard competitive models.

Finally, interactions across sources of monopsony power and the aforementioned macro propagation mechanisms may also be nonlinear. In Berger et al. (2024), we develop and estimate a model in which three sources of labor market monopsony interact: traditional (based on heterogeneous preferences of workers and search costs) and new (granular employers). Using estimates based on Norwegian data, we infer an average markdown of 21 percent, suggesting significant labor market power. Policy experiments reveal strong complementarities: removing search frictions narrows markdowns by 60 percent and significantly improves the allocation of workers to firms. In particular, many more workers reach the highest productivity firm at the top of the job ladder. Increasing competition in the bargaining process—shifting from duopsony to triopsony during wage renegotiations—narrows markdowns by 50 percent. Removing preference-based sources of monopsony narrows markdowns by 30 percent. That these numbers sum to greater than 100 percent indicates significant nonlinearities in how sources of monopsony power interact.

Implications of Large Firms for Micro Methods

When large firms affect the economy through granularity, strategic actions, and misallocation, then the microeconomic methods of analyzing labor market

power face fundamental challenges. The core problem is that the conventional micro methods rely on atomistic firm behavior, but granular firms create spillover effects that contaminate standard identification strategies.

As an example, suppose Firm A pays wages of $100 and after being the only firm in the market to receive a new patent, it expands its employment from 100 to 110 workers. How can we tell whether Firm A has labor market power? If Firm A was competitive, and thus faces a perfectly elastic supply curve of workers, it would expand its employment at the market wage—remember, the positive shock only affects Firm A. However, if Firm A has labor market power, then it faces an upward-sloping supply curve of potential workers, and expanding employment will require increasing wages, say to $110. In a micro empirical approach, Firm A's labor market power can be inferred from how much wages increase in order to increase employment, using other firms as a control group to estimate the treatment effect of the patent on wages and employment at Firm A.

This distinction between treatment and control fails in the setting of concentrated labor markets with a handful of dominant large employers. Suppose that the labor market has another large firm, Firm B, that also pays $100 and employs 100 workers. Both firms are granular. In equilibrium, Firm A's 10 new hires all come from Firm B. Comparing employment changes at the treated Firm A to the control Firm B delivers a swing in employment of 20 workers for a $10 wage increase. This contaminated estimate, which includes spillover effects to other firms, is what we call an *empirical elasticity*. It exceeds the true responsiveness of employment to wages at Firm A: an increase of 10 workers for a $10 wage increase. This true responsiveness, purged of spillovers, is what we call a *structural elasticity*. If Firm B responds by increasing its wages—that is, if Firm B takes a strategic action in response to Firm A's wage increase—the empirical elasticity becomes even more contaminated by general equilibrium effects and the picture gets murkier still.

What we will call the *macro approach* explicitly accounts for spillovers from treated firms to control firms that naturally occur in concentrated labor markets. By explicitly modeling the spillovers between Firm A and Firm B, researchers using the macro approach can recover the true responsiveness of employment to wages. This yields accurate estimates of the structural labor market power of firms, and accurate evaluations of policies that interact with labor market power.

Regression-Based Methods

The micro approach we just described is a difference-in-differences design, one of a family of regression-based methods that offer transparent identification and easy interpretation. The seminal Card and Krueger (1994) study of the effects of New Jersey's minimum wage on fast-food restaurant employment exemplifies this approach. Employment changes at treated establishments in New Jersey (which raised its minimum wage) are compared to a control group of establishments in Pennsylvania (which did not). This design has been enormously influential, particularly in the study of labor markets—for example, in comparing wages in local areas that have experienced a merger that led to an increase in labor market

concentration to those that had not (Arnold 2020); in comparing wages of workers with industry-specific skills in areas that had a substantial rise in labor market concentration from hospital mergers to areas that did not (Prager and Schmitt 2021); and in comparing how states that have seen a change in enforcement levels for noncompete agreements have lower wage increases than those that have not seen such a change (Johnson, Lavetti, and Lipsitz 2025).

However, these approaches rest on what researchers call the Stable Unit Treatment Value Assumption, which means that treating one group does not affect the control group. In fast food markets, what happens in New Jersey stays in New Jersey. This assumption holds under the traditional micro-based view of monopsony power in which firms are atomistic. But when labor markets are dominated by large employers, this assumption can fail in ways that create challenges for these research designs.

Consider three concrete ways the Stable Unit Treatment Value Assumption may fail in concentrated markets with dominant employers. First, large firms often operate across borders. When Walmart faces higher minimum wages in New Jersey, it might shift hours to Pennsylvania stores, contaminating the control group. Second, spillovers occur when large employers in Pennsylvania preemptively raise wages to retain workers who might otherwise cross the border to New Jersey for higher pay. Third, in other contexts, supply chain connections mean that when a major employer like Amazon adjusts wages in one state, its suppliers and competitors across state lines respond.

In short, markets with dominant employers—where economic policy would presumably have its largest effects—are also the settings that violate the Stable Unit Treatment Value Assumption.

Industrial Organization Methods

Common industrial organization methods measure labor market power by comparing a direct estimate of a worker's contribution to value-added to their current wage, as discussed earlier. But this computation of labor market power typically assumes that today's wage is all that workers and firms care about.

But this assumption may not hold true. In models by Jungerman (2023), Berger et al. (2024), and very recently Agostinelli et al. (2025), granular firms recognize how they influence not only current wages, but also the entire time path of compensation. Workers may accept lower current wages in exchange for training or higher future pay, making current wage comparisons misleading. Using data from Norway that allow them to separate the heterogeneous values that workers place on nonwage amenities, firm granularity, and search frictions, Berger et al. (2024) find that wage backloading accounts for 70 percent of markdowns measured from current wages. Under these alternative theories, the current wage at a granular firm is an inappropriate benchmark for computing labor market power.

The Macro Approach: A Symbiosis of Theory and Micro Estimates

Once researchers recognize the presence of granular employers, strategic actions, and their equilibrium spillovers, how can they proceed? The macro

approach offers one path forward by embedding the *empirical elasticities* we described earlier within general equilibrium models that account for spillovers and strategic interaction (Berger, Herkenhoff, and Mongey 2022a). This approach produces quantitative models that capture granularity, strategic actions, and misallocation, while being disciplined by micro data.

The macro approach consists of four steps. First, write down a theory of large firms in concentrated markets, capturing granular firms, strategic actions, and general equilibrium effects (as discussed earlier). Second, choose parameters of the model to ensure that it produces the same *empirical elasticities*, estimated from micro methods, when subjected to the same shock (say, it replicates the employment responses of firms in New Jersey following a large increase in the minimum wage). Third, because the *empirical elasticities* are contaminated by spillovers (as just noted), use the model to compute the sought-after *structural elasticities*, which are purged of spillovers. Fourth, use the framework to conduct counterfactuals that measure the positive and normative properties of alternative policies.

The key conceptual advance is in steps 2 and 3. The model can be disciplined by empirical elasticities, which are contaminated, and then produce structural elasticities, which are not. This approach produces a "big tent": micro empirical methods to inform macro model-based counterfactuals.

Amodio, Medina, and Morlacco (2025) provide a clear illustration. First, they write down a general equilibrium model where large manufacturing firms in concentrated local labor markets compete oligopsonistically for workers, while workers can choose between wage employment and self-employment. Second, they use the staggered rollout of Peru's rural electrification program for causal identification, allowing them to estimate how electrification differentially affects employment at firms with varying constraints in accessing electricity. They calibrate their model parameters so that when electrification occurs in the model, it replicates these empirical elasticities, while also matching cross-sectional correlations between market concentration and labor market outcomes. These correlations directly discipline the magnitude of general equilibrium spillovers by capturing how workers sort between sectors in response to market structure changes. Third, they use the model to recover structural elasticities that account for general equilibrium spillovers, including how electrification affects the sorting of workers between wage employment and self-employment, which changes the labor supply elasticity and amplifies or dampens the policy effects. Fourth, they use these structural elasticities to evaluate how labor market power affects the success of industrial development policies, finding that policies are only effective when they directly address the concentration of labor market power among large firms.

Of course, the macro approach typically requires stronger modeling assumptions than reduced-form micro evidence. However, it also provides an internally consistent way to measure labor market power among dominant employers and to evaluate policy counterfactuals. And it is precisely in markets with large firms and high concentration—where labor market power is most severe—that such macro tools are most informative and where policy interventions may be most needed. This makes the macro approach a valuable complement to existing micro methods.

Implications for Labor Market Policies

The macro approach gives a new perspective on policies designed to counteract firms' labor market power. In markets with large firms and substantial concentration, it qualifies the standard view of minimum wages as a primary tool against monopsony power. We then discuss more promising policies—such as antitrust enforcement and noncompete bans—that more directly target dominant employers in concentrated markets.

The Promise and Limitations of Minimum Wages

Minimum wages have long been viewed as the natural policy response to monopsony power. Joan Robinson's (1933) classic analysis showed that a monopsonist hiring below the competitive wage could be induced to raise wages to the competitive level through a binding minimum wage, increasing both wages and employment while eliminating the deadweight loss from labor market distortions. This insight has motivated decades of minimum wage advocacy as a remedy for labor market power.

However, this analysis assumes a single dominant employer, whereas in reality labor markets contain multiple firms of varying productivities. As we documented earlier, large firms have systematically larger markdowns. The presence of large firms in concentrated markets taking strategic actions fundamentally changes the picture, requiring a macro approach to understand minimum wage effects. Minimum wages now operate through three channels: direct effects on constrained firms, spillovers to unconstrained competitors, and reallocation from small to large establishments.

In Berger, Herkenhoff, and Mongey (2025), we ask: how well does a national minimum wage eliminate the deadweight loss from labor market power in the face of these realities? We develop a model that captures all three channels, while remaining consistent with well-identified empirical evidence from recent minimum wage increases in Seattle, Minneapolis, Brazil, and Germany (Jardim et al. 2022; Karabarbounis, Lise, and Nath 2022; Engbom and Moser 2022; Dustmann et al. 2022). Following the macro approach, we use these *empirical elasticities* to recover the *structural elasticities* governing monopsony power, then conduct general equilibrium policy analysis. We compute an optimal minimum wage of $8, generating welfare gains equivalent to 0.2 percent of lifetime consumption.

These modest gains emerge entirely through channels absent in traditional models. Granularity drives a beneficial reallocation mechanism: higher minimum wages constrain small, low-productivity firms, inducing workers to move toward large, high-productivity employers. However, the optimal minimum wage remains modest because spillovers—while present—are not strong enough to substantially reduce markdowns at the biggest firms, which have the most market power and generate the largest deadweight losses. Raising the minimum wage too high rations workers out of employment at small firms without meaningfully constraining the dominant employers responsible for most welfare losses. Under a counterfactual assumption that firms are atomistic, the optimal minimum wage collapses to just

$0.70, with essentially zero welfare gains because the reallocation channel shuts down entirely.

Our macro approach also explains why micro empirical studies of different labor markets often find conflicting results. As we discussed earlier, there is considerable heterogeneity across the concentration of labor markets. In our model, minimum wages are more likely to generate positive employment effects via general equilibrium spillovers in the labor markets that are concentrated. However, in the unconcentrated markets that are closer to the textbook theory of competitive markets, minimum wages primarily generate negative employment effects. This is consistent with empirical evidence (for example, Azar et al. 2024; Popp 2024). The heterogeneity in empirical findings about minimum wages thus reflects real differences in market structure rather than methodological inconsistencies or identification failures.

The macro approach also reveals fundamental limitations of minimum wage policy for addressing labor market power. Recall that 40 percent of wage payments occur in highly concentrated markets where large firms dominate, as shown earlier in Figure 1. Because the large firms that have the most labor market power typically pay above minimum wage, broad-based minimum wage policies miss the core of the problem facing low-wage workers. The minimum wage has no direct effect on the large firms in concentrated labor markets, which are the firms responsible for welfare losses under the macro approach. In our model, an optimally chosen minimum wage captures only about 3 percent of the potential efficiency gains available from eliminating labor market power entirely. These calculations highlight that, while such a minimum wage raises welfare, it leaves most of the efficiency losses from monopsony power untouched, especially in concentrated markets dominated by large firms.

Noncompete Clauses and Antitrust

Unlike minimum wages that primarily constrain smaller firms with limited labor market power, several policies directly target dominant employers.

One such policy would be to ban firms from including noncompete clauses in employment contracts. Between 36 and 60 million US workers are bound by these agreements (Colvin and Shierholz 2019). Using a macro approach with granular firms and search frictions, Gottfries and Jarosch (2025a) find that banning noncompetes raises wages by 4 percent. However, the direct effects—in this case, the higher wages of individual workers whose noncompetes are lifted—are small. However, the *spillover effects* of higher wages of all workers as competition increases are large, highlighting the importance of modeling macro effects. (In this symposium, the essay by Evan Starr focuses on the economics of noncompete agreements.)

Another possible policy is greater antitrust enforcement in labor markets. By modeling granularity explicitly, the macro approach can be used to study how mergers of large firms are likely to affect wages and employment. In Berger et al. (2025), we adopt this approach and develop a macro model of multi-plant ownership and mergers. We estimate the model on US Census data and show that

it reproduces the effects of actual mergers on workers and firms. We then show that stricter antitrust enforcement in labor markets yields large welfare gains for households and mitigates labor market power. This is consistent with the recent direction of US policy; specifically, the 2023 Merger Guidelines promulgated by the Federal Trade Commission and the US Department of Justice explicitly states that the agencies will consider worker harm on a stand-alone basis as a rationale for blocking proposed mergers. (The paper in this symposium by Elena Prager delves more deeply into labor market antitrust issues.)

Further Extensions and Open Questions

The key insight here is that micro empirical methods measure *direct effects*, but noncompete labor agreements and mergers can lead to aggregate harms for workers that extend far beyond workers directly bound by such agreements. While new models of large firms in concentrated labor markets are a useful starting point, further advances can help to understand trade-offs that these models have so far omitted.

As one example, the existing macro models of labor market power do not weigh welfare gains from reduced labor market power against the potential for noncompete bans and stricter antitrust laws to reduce innovation, investment, research and development, and growth (for example, Cavenaile, Celik, and Tian 2021; Phillips and Zhdanov 2024). Some promising work links labor market power that firms have over narrowly specialized inventors to growth and innovation. For example, incumbent granular firms may engage in defensive hiring of researchers to hinder new competition (Fernández-Villaverde, Yu, and Zanetti 2025; Babalievsky 2022). Other key aspects of labor market power are not yet incorporated into macro models either. What if incumbent granular firms manipulate or control the conditions for occupation licensing, or lobby for nonenforcement of noncompetes?

An overall message emerging from this literature is that because labor market power is concentrated among the largest firms in concentrated labor markets, the most effective interventions in this particular setting will seek to target the source of the largest wage markdowns directly, rather than using broad-based policies that potentially target competitive, low markdown firms as well.

Conclusion

Evidence for granular firms with monopsonistic labor market power is now substantial. However, the possible ways in which dominant firms shape aggregate labor market outcomes are far from fully explored. How do wage-setting decisions by major employers cascade through regional labor markets? What are the innovation consequences when a handful of firms control access to top talent? How does monopsony power by large retailers affect the allocation of production across supply chains? These questions put dominant firms at the heart of both labor economics and macroeconomic policy.

Our article puts forth three main arguments. First, the presence of large firms in concentrated markets poses challenges for micro approaches to measure labor market power, as these approaches typically rely on the assumption of atomistic firm behavior. Second, once one recognizes the presence of dominant employers, three interconnected mechanisms drive aggregate effects: granularity makes individual firm actions matter for entire markets, which enables strategic interactions that amplify market power, which in turn generates misallocation as productive firms restrict hiring. These mechanisms can reinforce each other. Large firms use their granular position to act strategically, and this strategic behavior pushes workers toward less productive employers. Third, the macro approach of directly modeling and estimating the strength of these mechanisms fundamentally alters policy advice and welfare losses from market power. Because labor market power is concentrated among the largest firms in concentrated markets, effective interventions must target the source of the largest wage markdowns directly rather than broad-based policies that treat all firms equally.

Understanding how dominant firms shape aggregate labor market outcomes is not just an academic exercise. Instead, it is essential for designing labor market policy that puts workers and firms on a more equal footing. The welfare and output losses documented across the research on labor market power are comparable to estimates of those from product market distortions. (As a product market example, Edmond, Midrigan, and Xu (2023) estimate welfare costs of heterogeneous firms and endogenous markups, estimated with US Census of Manufactures data.) The similar magnitude of welfare losses from labor and product market power implies that, for researchers, policymakers, and antitrust authorities, addressing employers' market power in the labor market should be a priority comparable to the study of imperfect competition in product markets.

■ *We thank Timothy Taylor for detailed editoral comments. We also thank Gabe Chodorow-Reich, Pat Kline, Jeffrey Kling, Jonathan Parker, Ben Schoefer, Isaac Sorkin, and Daniel Xu for helpful comments on an earlier draft.*

References

Agostinelli, Francesco, Domenico Ferraro, Giuseppe Sorrenti, and Leonard Treuren. 2025. "Employment Relationships, Wage Setting, and Labor Market Power." NBER Working Paper 34439.

Amodio, Francesco, Pamela Medina, and Monica Morlacco. 2025. "Labor Market Power, Self-Employment, and Development." *American Economic Review* 115 (9): 3014–57.

Arnold, David. 2020. "Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.3476369.

Ashenfelter, Orley, and Štěpán Jurajda. 2025. "McMonopsony." Unpublished.

Azar, José A., Steven T. Berry, and Ioana Marinescu. 2022. "Estimating Labor Market Power." NBER Working Paper 30365.

Azar, José, Emiliano Huet-Vaughn, Ioana Marinescu, Bledi Taska, and Till von Wachter. 2024. "Minimum Wage Employment Effects and Labor Market Concentration." *Review of Economic Studies* 91 (4): 1843–83.

Azar, José, Ioana Marinescu, and Marshall Steinbaum. 2022. "Labor Market Concentration." *Journal of Human Resources* 57 (S), S167–99.

Azkarate-Askasua, Miren, and Miguel Zerecero. 2025. "Union and Firm Labor Market Power." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4323492.

Babalievsky, Fil. 2022. "Misallocation in the Market for Inventors." Unpublished.

Bagga, Sadhika. 2023. "Firm Market Power, Worker Mobility, and Wages in the US Labor Market." *Journal of Labor Economics* 41 (S1): S205–56.

Belot, Michèle, Philipp Kircher, and Paul Muller. 2022. "How Wage Announcements Affect Job Search—A Field Experiment." *American Economic Journal: Macroeconomics* 14 (4): 1–67.

Berger, David, Thomas Hasenzagl, Kyle Herkenhoff, Simon Mongey, and Eric A. Posner. 2025. "Merger Guidelines for the Labor Market." *Journal of Monetary Economics* 153: 103785.

Berger, David, Kyle Herkenhoff, Andreas R. Kostøl, and Simon Mongey. 2024. "An Anatomy of Monopsony: Search Frictions, Amenities, and Bargaining in Concentrated Markets." In *NBER Macroeconomics Annual*, Vol. 38, edited by Martin Eichenbaum, Erik Hurst, and Valerie Ramey, 1–47. University of Chicago Press.

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2022a. "Labor Market Power." *American Economic Review* 112 (4): 1147–93.

Berger, David, Kyle F. Herkenhoff, and Simon Mongey. 2022b. "Minimum Wages, Efficiency and Welfare." NBER Working Paper 29662.

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2025. "Minimum Wages, Efficiency, and Welfare." *Econometrica* 93 (1): 265–301.

Bhaskar, V., Alan Manning, and Ted To. 2002. "Oligopsony and Monopsonistic Competition in Labor Markets." *Journal of Economic Perspectives* 16 (2): 155–74.

Bils, Mark, Bariş Kaymak, and Kai-Jie Wu. 2025. "Robinson Meets Roy: Monopsony Power and Comparative Advantage." NBER Working Paper 34165.

Bloesch, Justin, Birthe Larsen, and Anders Yding. 2025. "Monopsony with Recruiting." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.5020154.

Boal, William M., and Michael R. Ransom. 1997. "Monopsony in the Labor Market." *Journal of Economic Literature* 35 (1): 86–112.

Brooks, Wyatt J., Joseph P. Kaboski, Yao Amber Li, and Wei Qian. 2019. "Exploitation of Labor? Classical Monopsony Power and Labor's Share," NBER Working Paper 25660.

Brooks, Wyatt J., Joseph P. Kaboski, Yao Amber Li, and Wei Qian. 2021. "Exploitation of Labor? Classical Monopsony Power and Labor's Share." *Journal of Development Economics* 150: 102627.

Burdett, Kenneth, and Kenneth L. Judd. 1983. "Equilibrium Price Dispersion." *Econometrica* 51 (4): 955–69.

Burdett, Kenneth, and Dake T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review* 39 (2): 257–73.

Caldwell, Sydnee, Ingrid Haegele, and Jörg Heining. 2025. "Bargaining and Inequality in the Labor Market." NBER Working Paper 33396.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *Quarterly Journal of Economics* 128 (3): 967–1015.

Card, David, and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84 (4): 772–93.

Carlsson, Mikael, Julián Messina, and Oskar Nordström Skans. 2021. "Firm-Level Shocks and Labour Flows." *Economic Journal* 131 (634): 598–623.

Cavenaile, Laurent, Murat Alp Celik, and Xu Tian. 2021. "The Dynamic Effects of Antitrust Policy on Growth and Welfare." *Journal of Monetary Economics* 121: 42–59.

Chan, Mons, Elena Mattana, Sergio Salgado, and Ming Xu. 2023. "Wage Setting and Passthrough: The Role of Market Power, Production Technology and Adjustment Costs." http://dx.doi.org/10.2139/ssrn.4544524.

Cheremukhin, Anton, and Paulina Restrepo-Echavarria. 2025. "An Information-Based Theory of Monopsony Power." Federal Reserve Bank of Dallas Working Paper 2518.

Cheremukhin, Anton, Paulina Restrepo-Echavarria, and Antonella Tutino. 2020. "Targeted Search in Matching Markets." *Journal of Economic Theory* 185: 104956.

Clemens, Jeffrey, and Michael R. Strain. 2021. "The Heterogeneous Effects of Large and Small Minimum Wage Changes: Evidence over the Short and Medium Run Using a Pre-analysis Plan." NBER Working Paper 29264.

Colvin, Alexander J. S., and Heidi Shierholz. 2019. *Noncompete Agreements: Ubiquitous, Harmful to Wages and to Competition, and Part of a Growing Trend of Employers Requiring Workers to Sign Away Their Rights.* Economic Policy Institute Report.

Derenoncourt, Ellora, Clemens Noelke, David Weil, and Bledi Taska. 2021. "Spillover Effects from Voluntary Employer Minimum Wages." NBER Working Paper 29425.

Dobbelaere, Sabien, Jordi Jaumandreu, and Jacques Mairesse. 2025. "Market Power and Technology in US Manufacturing." Unpublished.

Dobbelaere, Sabien, and Jacques Mairesse. 2013. "Panel Data Estimates of the Production Function and Product and Labor Market Imperfections." *Journal of Applied Econometrics* 28 (1): 1–46.

Dube, Arindrajit, Jeff Jacobs, Suresh Naidu, and Siddharth Suri. 2020. "Monopsony in Online Labor Markets." *American Economic Review: Insights* 2 (1): 33–46.

Dube, Arindrajit, Suresh Naidu, and Adam D. Reich. 2022. "Power and Dignity in the Low-Wage Labor Market: Theory and Evidence from Wal-Mart Workers." NBER Working Paper 30441.

Dustmann, Christian, Attila Lindner, Uta Schönberg, Matthias Umkehrer, and Philipp vom Berge. 2022. "Reallocation Effects of the Minimum Wage." *Quarterly Journal of Economics* 137 (1): 267–328.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu. 2023. "How Costly Are Markups?" *Journal of Political Economy* 131 (7): 1619–75.

Engbom, Niklas, and Christian Moser. 2022. "Earnings Inequality and the Minimum Wage: Evidence from Brazil." *American Economic Review* 112 (12): 3803–47.

Esquierro, Leon, and Sergio da Silva. 2024. "Is the Brazilian Labor Market Granular?" *Economics Bulletin* 44 (2): 576–85.

Faberman, R. Jason, Andreas I. Mueller, Ayşegül Şahin, and Giorgio Topa. 2022. "Job Search Behavior among the Employed and Non-employed." *Econometrica* 90 (4): 1743–79.

Fernández-Villaverde, Jesús, Yang Yu, and Francesco Zanetti. 2025. "Defensive Hiring and Creative Destruction." NBER Working Paper 33588.

Gabaix, Xavier. 2011. "The Granular Origins of Aggregate Fluctuations." *Econometrica* 79 (3): 733–72.

Gottfries, Axel, and Gregor Jarosch. 2025a. "Dynamic Monopsony with Granular Firms." NBER Working Paper 31965.

Gottfries, Axel, and Gregor Jarosch. 2025b. "Wage Fixing." NBER Working Paper 33501.

Hall, Robert E., and Alan B. Krueger. 2012. "Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-the-Job Search." *American Economic Journal: Macroeconomics* 4 (4): 56–67.

Hasenzagl, Thomas. 2024. "Scaling Up: How Technology and Policy Shape Firm Dynamics." Unpublished.

Holmes, Thomas J., and James A. Schmitz, Jr.. 1994. "Resistance to Technology and Trade between Areas." Federal Reserve Bank of Minneapolis Research Department Staff Report 184.

Jäger, Simon, Christopher Roth, Nina Roussille, and Benjamin Schoefer. 2024. "Worker Beliefs about Outside Options." *Quarterly Journal of Economics* 139 (3): 1505–56.

Jardim, Ekaterina, Mark C. Long, Robert Plotnick, Emma van Inwegen, Jacob Vigdor, and Hilary Wething. 2022. "Minimum-Wage Increases and Low-Wage Employment: Evidence from Seattle." *American Economic Journal: Economic Policy* 14 (2): 263–314.

Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin. 2024. "Granular Search, Market Structure, and Wages." *Review of Economic Studies* 91 (6): 3569–607.

Johnson, Matthew S., Kurt Lavetti, and Michael Lipsitz. 2025. "The Labor Market Effects of Legal Restrictions on Worker Mobility." *Journal of Political Economy* 133 (9): 2735–93.

Jungerman, William. 2023. "Dynamic Monopsony and Human Capital." Unpublished.

Karabarbounis, Loukas, Jeremy Lise, and Anusha Nath. 2022. "Minimum Wages and Labor Markets in the Twin Cities." NBER Working Paper 30239.

Kleiner, Morris M., and Evan J. Soltas. 2023. "A Welfare Analysis of Occupational Licensing in US States." *Review of Economic Studies* 90 (5): 2481–516.

Kline, Patrick, Neviana Petkova, Heidi Williams, and Owen Zidar. 2019. "Who Profits from Patents? Rent-Sharing at Innovative Firms." *Quarterly Journal of Economics* 134 (3): 1343–404.

Kovalenko, Tim, Claus Schnabel, and Heiko Stüber. 2022. "Is the German Labour Market Granular?" *Applied Economics Letters* 29 (1): 41–48.

Kroft, Kory, Yao Luo, Magne Mogstad, and Bradley Setzler. 2025. "Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry." *American Economic Review* 115 (9): 2926–69

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.

Lehr, Nils H. 2023. "Does Monopsony Matter for Innovation?" Unpublished.

Manning, Alan. 2003. *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press.

Mortensen, Dale T., and Christopher A. Pissarides. 1999. "New Developments in Models of Search in the Labor Market." In *Handbook of Labor Economics*, Vol. 3, edited by Orley C. Ashenfelter and David Card, 2567–627. North-Holland.

Neumark, David, and Peter Shirley. 2022. "Myth or Measurement: What Does the New Minimum Wage Research Say about Minimum Wages and Job Loss in the United States?" *Industrial Relations* 61 (4): 384–417.

Phillips, Gordon M., and Alexei Zhdanov. 2013. "R&D and the Incentives from Merger and Acquisition Activity." *Review of Financial Studies* 26 (1): 34–78.

Phillips, Gordon M., and Alexei Zhdanov. 2024. "Venture Capital Investments, Merger Activity, and Competition Laws around the World." *Review of Corporate Finance Studies* 13 (2): 303–34.

Popp, Martin. 2024. "Minimum Wages in Concentrated Labor Markets." IZA Discussion Paper 17357.

Posner, Eric A. 2021. *How Antitrust Failed Workers*. Oxford University Press.

Posner, Eric A., and Ruth Zheng. 2025. "The Silicon Valley No-Poach Conspiracy." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.5245922.

Prager, Elena, and Matt Schmitt. 2021. "Employer Consolidation and Wages: Evidence from Hospitals." *American Economic Review* 111 (2): 397–427.

Robinson, Joan. 1933. *The Economics of Imperfect Competition*. Macmillan.

Roussille, Nina, and Benjamin Scuderi. 2023. "Bidding for Talent: A Test of Conduct in a High-Wage Labor Market." NBER Working Paper 33848.

Rubens, Michael, Yingjie Wu, and Mingzhi Xu. 2025. "Estimating Factor Price Markdowns Using Production Models." *International Journal of Industrial Organization* 102: 103177.

Schmitz, James A., Jr. 2020. "Monopolies Inflict Great Harm on Low- and Middle-Income Americans." Federal Reserve Bank of Minneapolis Staff Report 601.

Schmitz, James A., Jr. 2005. "What Determines Productivity? Lessons from the Dramatic Recovery of the US and Canadian Iron Ore Industries Following Their Early 1980s Crisis." *Journal of Political Economy* 113 (3): 582–625.

Seegmiller, Bryan. 2021. "Valuing Labor Market Power: The Role of Productivity Advantages." Unpublished.

Shi, Liyan. 2023. "Optimal Regulation of Noncompete Contracts." *Econometrica* 91 (2): 425–63.

Staiger, Douglas O., Joanne Spetz, and Ciaran S. Phibbs. 2010. "Is There Monopsony in the Labor Market? Evidence from a Natural Experiment." *Journal of Labor Economics* 28 (2): 211–36.

Starr, Evan P., J. J. Prescott, and Norman D. Bishara. 2021. "Noncompete Agreements in the US Labor Force." *Journal of Law and Economics* 64 (1): 53–84.

Starr, Evan, and Robert H. Smith. 2019. *The Use, Abuse, and Enforceability of Noncompete and No-Poach Agreements*. Economic Innovation Group Report.

Trottner, Fabian. 2025. "A Matter of Taste: A Unified Approach to Modeling Monopsony." Unpublished.

US Bureau of Labor Statistics. 2025. *Graphics for Economic News Releases: Employment by Size of Establishment, Private Industry*. https://www.bls.gov/charts/county-employment-and-wages/employment-by-size.htm (accessed December 23, 2025).

USDA Economic Research Service. 2025. "Commuting Zones and Labor Market Areas." September 4. https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas.

Volpe, Oscar. 2024. "Job Preferences, Labor Market Power, and Inequality." Unpublished.

Webber, Douglas A. 2016. "Firm-Level Monopsony and the Gender Pay Gap." *Industrial Relations* 55 (2): 323–45.

Yeh, Chen, Claudia Macaluso, and Brad Hershbein. 2022. "Monopsony in the US Labor Market." *American Economic Review* 112 (7): 2099–138.

Journal of Economic Perspectives—Volume 40, Number 1—Winter 2026—Pages 115–138

Antitrust Enforcement in Labor Markets

Elena Prager

In 2021, the world's largest book publisher, Penguin Random House, proposed buying another of the "big five" US book publishers, Simon & Schuster. The US Department of Justice sued to block the merger on the grounds that it would harm competition. The US District Court for the District of Columbia agreed, blocking the merger in 2022 (*US v. Bertelsmann 2022*). The case was extraordinary because the primary argument for blocking the merger was not the conventional concern that the merged firm would gain monopoly power and raise prices for the books sold to readers. Instead, the court agreed with the concern of the Department of Justice that the merged firm would gain *monopsony* power that could "substantially lessen competition to acquire 'the publishing rights to anticipated top-selling books.'" The book publishing case thus became the first US merger case to be brought—and won—primarily on the grounds of harm to labor market competition. Such a case would likely have been unthinkable a mere decade earlier.

For many decades, the US antitrust laws that exist to protect competitive markets and the welfare gains that flow from competition were almost exclusively applied to markets for output. They were almost never applied to markets for labor. Proposed corporate mergers were not evaluated on whether they were likely, say, to give the merged firm power to suppress compensation for its employees by absorbing a key competitor. Consider, for instance, a merger between two hospitals that together dominate a local labor market for nurses. Antitrust authorities would

■ *Elena Prager is Assistant Professor of Economics, University of Rochester Simon Business School, Rochester, New York. She is also Faculty Research Fellow, National Bureau of Economic Research, Cambridge, Massachusetts. Her email address is eprager@simon.rochester.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20241446.

have considered whether such a merger might raise prices for patients and insurers, but not whether it would reduce competition for hiring nurses, potentially resulting in lower wages or worsened working conditions. In the last decade or so, antitrust authorities have begun to take labor market power more seriously, as have many academic researchers. A loose consensus has emerged that labor market concentration and related sources of labor market power can meaningfully harm labor market competition. But controversy remains on whether antitrust tools are actually capable of protecting competition in labor markets. Until recently, there was little empirical evidence on the consequences of the types of labor market power that are actionable under antitrust law. The literature on labor market power historically focused on search frictions, which are typically not actionable through the tools of antitrust. That evidence gap is rapidly being filled. As discussed below, the evidence shows that harms to labor market competition can indeed arise through mechanisms that antitrust enforcement can tackle.

Antitrust enforcement activity has begun to reflect antitrust law's preexisting, but previously dormant, jurisdiction over labor market competition. The most important antitrust enforcers in the United States are the Federal Trade Commission (FTC) and the Antitrust Division of the Department of Justice (DOJ). Across various areas of antitrust enforcement, the number of FTC and DOJ actions with a labor market component has risen from one in 2010–2014 to seven in 2020–2024 (Kariel, Schneebacher, and Walker 2024). To put these low counts in perspective, consider that the FTC and DOJ collectively bring only 25 to 30 merger cases in a typical year (Thurman Arnold Project 2020).

The uptick in labor market enforcement activity is mirrored in policy documents, staffing decisions, and court cases. For example, the Federal Trade Commission and the Department of Justice have been jointly issuing the Merger Guidelines since 1968, with subsequent revisions once or twice a decade. For over 50 years, the Merger Guidelines have included a discussion of input markets, but without mentioning labor inputs, despite labor's outsized importance to production. The most recent 2023 revision of the Merger Guidelines is the first to call out labor markets, stating explicitly that the agencies will consider labor market impacts in merger evaluation (DOJ and FTC 2023). Agency staffing decisions also reflect a new focus on labor markets. In 2022, for the first time, the DOJ brought in a labor economist to serve in a high-level antitrust position. In 2025, the FTC launched a new task force on labor market competition.

The aim of this essay is to explain how antitrust regulation is beginning to be used in labor markets, the evidence for and against its use, and the remaining evidence gaps standing in the way of more effective use. I begin with a brief primer on antitrust law and the recent sea change in its application to labor markets. The following section describes the key economic theories of labor market competition and the state of the empirical evidence, as they relate to antitrust enforcement. Finally, I turn to the practical issues that arise when applying antitrust tools to labor markets and some key open questions where additional research is needed to inform enforcement.

The History of Antitrust (Non)Enforcement in Labor Markets

This section reviews the recent evolution in antitrust enforcement's approach to labor market power, first in a US context and then with a brief look at Europe.

US Antitrust Enforcement

The United States has two keystone pieces of antitrust legislation: the Sherman Act of 1890 and the Clayton Act of 1914. Together, these acts prohibit collusion between firms: monopolization, which is defined as the gaining or maintenance of monopoly power through anticompetitive (loosely speaking, "unnatural") means such as predatory pricing; and mergers that would harm competition. The rationale for these prohibitions is that increases in monopoly power and behaviors like collusion prevent society from reaping the full benefits of competition.

Notably, the mere possession of monopoly power is not prohibited by antitrust law. Enforcers do not have authority over high *levels* of concentration and cannot easily break up large firms or undo consummated mergers.[1] The hope is that when a market is characterized by monopoly power, the market will self-correct by attracting new entrants that seek a share of the monopoly profit. This reliance on the efficacy of entry explains why barriers to entry also fall under the purview of antitrust enforcement. In practice, evidence is mixed on how effectively entry disciplines market power. For example, in a classic paper using data on entry, growth, and exit of US manufacturing firms during 1963–1982, Dunne, Roberts, and Samuelson (1988) find that most entrants do not survive for long. On the other side, Starc and Wollmann (2025) find that entry in the generic drug manufacturing industry dissipates the harms from collusion.

Many areas of antitrust enforcement rely on economics. Certain conduct is *per se* illegal, which means that it is unlawful regardless of the magnitude of its effects. Examples include blatant price fixing and market allocation. In such cases, antitrust enforcement makes little use of economics because the law is unambiguous. However, much of antitrust law concerns outcomes that are theoretically or legally ambiguous. For example, a corporate merger may or may not be anticompetitive, depending on many factors including how closely the merging firms currently compete. An "exclusive dealing" contract in which one firm agrees to buy a product from only one supplier, or a supplier agrees to sell to only one buyer, may be anticompetitive if its primary effect is to erect barriers to entry, or it may be procompetitive if it resolves principal-agent problems between a supplier and a distributor. Adjudication of such ambiguous cases relies heavily on economics, including sophisticated modeling and data analyses, to determine the effects of competition. However, there are also

[1] The most notable exceptions are the forced breakups of Standard Oil in 1911 and of AT&T in 1984. As of this writing, the antitrust case against Alphabet includes discussions of breaking up firms under arguments of monopolization; however, courts rejected similar claims by the government in its case against Meta in November 2025.

situations in which antitrust law diverges from the underlying economics, such as the treatment of tacit collusion (a subject to which I will return).

Economics is used particularly heavily in the evaluation of corporate mergers. The Hart-Scott-Rodino (HSR) Antitrust Improvements Act of 1976 requires that firms planning a merger or acquisition first notify the federal government of intended transactions above a certain size (in 2025, the threshold was $126.4 million). The firms must also provide basic information about their competitive positions and product portfolios. This information is reviewed by the Federal Trade Commission or the Department of Justice, depending on the industry and other details of the proposed transaction. If further investigation seems warranted, the reviewing antitrust agency obtains additional data, documents, and testimony from the merging parties and other industry participants. If this information offers a *prima facie* case that the merger might be harmful, then more sophisticated economic analysis is often performed before challenging a merger. Few proposed mergers reach this stage: of the thousands of intended and completed mergers in the United States each year, only a few dozen are investigated in depth, and only about half of those are challenged in court (Thurman Arnold Project 2020).

In principle, antitrust law applies to labor markets, because the law is agnostic as to market type. The text of the law does not privilege competition in product markets over competition in labor or other input markets.[2] Until recently, however, the cases chosen for enforcement by the Federal Trade Commission and the Department of Justice almost never involved protecting labor market competition. No corporate merger case had ever been challenged primarily on grounds of harm to labor market competition until the 2021 proposed merger of the book publishers Penguin Random House and Simon & Schuster. Only a handful of labor market collusion cases were brought over the century following the passage of the Sherman and Clayton Acts. In fact, in the first few decades after the Sherman Act of 1890, antitrust law was used in support of employers against labor unions, with unions viewed as cartels. Clarifications in the Clayton Act of 1914 and the Norris-Laguardia Act of 1932 eventually put this practice to rest.

As discussed in the introduction, antitrust enforcement has recently begun to be used to protect labor market competition. However, using antitrust enforcement in this way creates a potential tension: What if the higher wages from stronger labor market competition are passed through to consumers in the form of higher prices? Existing US antitrust law is clear on this point. Mergers that substantially harm competition in at least one market are illegal, even if they are procompetitive in other markets. Specifically, Section 7 of the Clayton Act prohibits mergers that substantially lessen competition "in any line of commerce in any section of the country." To put it another way, a merger that benefits consumers by reducing prices could be illegal—if it does so through anticompetitive reductions in wages (or other input costs).

[2] Section 8 of the Clayton Act is explicitly about output market competition. This is the exception rather than the rule.

The public discourse around this point has been muddled by a misunderstanding of the welfare standard used in merger law. The United States uses the so-called "consumer welfare" standard, which is often erroneously interpreted as meaning that only harms to final consumers are a valid reason to block a merger. In actuality, the consumer welfare standard should be understood to mean the welfare of trading partners opposite the merging parties (Kaplow 2024). In many merger cases, those trading partners are indeed consumers. In a merger of firms that draw upon the same labor pool, the parties opposite the merger may be workers. The existing consumer welfare standard should therefore be understood as consistent with concerns about the effects of mergers on competition in labor markets. Moreover, it is not obvious that a merger that reduces labor market competition will generally benefit consumers. I return to this point below in the section on merger enforcement.

Even as US antitrust authorities are beginning to include labor market competition in evaluations of proposed mergers, they have also brought several cases against alleged illegal suppression of competition for workers.[3] These cases, which all involved allegations of wage-fixing or labor market allocation, have spanned a range of industries including tech (the Silicon Valley no-poaching case); no-poaching and wage-fixing cases in health care (the *DaVita*, *Manahe*, *Jindal*, *VDA*, and *Lopez* cases); aerospace (Raytheon); poultry processing (*Cargill*); and e-sports (*Activision Blizzard*).[4] For an excellent summary of recent labor market cases pursued by the Department of Justice, see Athey et al. (2023, 2024).

Some of these cases involved egregious actions by firms, such as text message trails stating "We all have a mutual agreement [. . .] stay within the same hourly rate" for worker pay (*U.S. v. Lopez 2023*). However, taken as a group, the antitrust cases focused on no-poaching and wage-fixing have had mixed success in court. The

[3]Additional conduct cases have been brought by private plaintiffs, including in education and sports. For example, a case against Duke University and the University of North Carolina for a no-poaching agreement that suppressed faculty wages was settled for a payment of $73 million. A series of lawsuits against the National Collegiate Athletics Association (NCAA) in recent years has led to players being able to receive payments from sponsors and from universities directly, as well as to transfer between schools more easily. There is currently a lawsuit brought by several professional tennis players and the Professional Tennis Players Association against the Association of Tennis Professionals (ATP) that governs men's tennis and the Women's Tennis Association (WTA) that governs women's tennis that they operate as a cartel and suppress players' earnings. Private litigation plays an important role in antitrust enforcement, but is beyond the scope of this article.

[4]The key allegations in these cases were as follows: In the Silicon Valley case, Apple, Google, and several other tech firms were accused of allocating the market for labor by agreeing not to recruit one another's employees (*In re: High-Tech Employee Antitrust Litigation* (2011) and additional court cases). DaVita is a dialysis company accused of allocating the market for certain employees (*US v. DaVita Inc.* 2021). The *Manahe*, *Jindal*, *VDA*, and *Lopez* cases all involved accusations of fixing wages for various health care professions: home health aides (*US v. Manahe* 2023), physical therapists (*US v. Jindal* 2020), and nurses (*US v. VDA* 2021; *US v. Lopez* 2023). Raytheon is an aerospace company accused with others of an alleged attempt to limit wages and mobility of aerospace engineers (*U.S. v. Patel* 2021). Cargill, Hormel, and other poultry processing companies were accused of sharing wage information in an effort to hold down wages (*US v. Cargill* 2022). Activision Blizzard is a video game company accused of illegally limiting compensation to professional e-sports players (*US v. Activision Blizzard* 2023).

federal government's antitrust agencies lost the Raytheon no-poaching case, despite the courts' recognition that labor market allocation took place (Posner 2023), and lost several of the health care cases (the *DaVita*, *Manahe*, and *Jindal* cases). The government obtained a guilty plea in the *VDA* case, settlements in the Silicon Valley and *Cargill* cases, and a consent decree in the *Activision Blizzard* case. The government obtained its first criminal conviction for wage-fixing in the *Lopez* case. The *Lopez* conviction was viewed as an important win for the agencies, because criminal liability is generally viewed as a stronger deterrent than civil liability.

US law is precedent-based, meaning that these early successes and failures of antitrust enforcement in labor markets will affect future enforcement. In some other countries, certain antitrust cases are handled administratively, with cases evaluated by specialized technical experts. By contrast, a judge hearing an antitrust case in the United States typically has limited experience with antitrust cases and no formal training in economics, which makes US enforcement particularly reliant on arguments and methods that are already accepted by the courts as precedent. The US legal system can therefore be slow to change in response to new methods and evidence, which makes it difficult to bring cases on novel grounds, such as labor market harms.

Other Jurisdictions

Interest in labor market antitrust enforcement has also been increasing outside the United States in recent years, with European jurisdictions playing an outsized role. The European Union's Directorate General for Competition (DG COMP) and the United Kingdom's Competition and Markets Authority (CMA) have organized workshops on labor market competition, and the CMA published a first-of-its-kind report on the state of labor market competition in 2024 (CMA Microeconomics Unit 2024). The pace of labor market–related cases and actions has risen dramatically across the globe, though the absolute counts remain low, rising from 3 in 2010–2014 to 19 in 2020–2024 (Kariel, Schneebacher, and Walker 2024). Kariel, Schneebacher, and Walker (2024) provide a more comprehensive overview of recent developments outside the United States. As in the United States, enforcers have recently acted on labor market collusion in a wide range of industries, including sports, retail, health care, and beauty services.

Naturally, the details of antitrust law and the legal system vary by jurisdiction. The most important departures from the US system are in merger enforcement. In both the United Kingdom and European Union, a large fraction of merger enforcement is conducted administratively, outside the general court system. The United Kingdom's CMA reviews and makes decisions on mergers internally, without arguing before a court. The European Union's DG COMP issues its own decisions as well, although they are subject to review by a general court. These structures expand the scope for novel economic analyses relative to US merger cases. Of course, European antitrust enforcement does not set a formal legal precedent for US regulators and courts. However, there is a longstanding culture of cooperation among the key antitrust authorities, such that antitrust ideas often diffuse across jurisdictions. Indeed, antitrust law in other large economies is broadly similar to the United States.

Evidence on Actionable Labor Market Power

Are the actions available to antitrust enforcers actually capable of protecting competition in labor markets? The answer is not immediately obvious. Several threats to competition are not actionable to enforcers. Market power is only actionable by antitrust enforcers if it is being abused in specific ways, or if it is being increased through collusion or a proposed merger.

Theoretical models propose two distinct classes of explanations for how labor market power might arise. The first class, search-and-matching models, shows how upward-sloping residual labor supply curves can arise from natural market frictions such as job search costs. In these models, firms are typically infinitesimal, leaving no role for collusion, mergers, or other behavior that would be actionable under antitrust law. The second class of models takes seriously that firms can be large enough for meaningful strategic interactions. In turn, strategic interactions and concentration create a role for antitrust enforcement.

Until recently, however, there was little empirical evidence on this second class of explanations. This section first reviews both classes of models, and then turns to the recent empirical evidence. The recent research finds that harms to labor market competition can indeed arise through mechanisms that are actionable under antitrust law. These research findings have plainly played a part in the direction of enforcement. For example, during the writing of the 2023 Merger Guidelines, senior federal officials appealed to specific research papers in explaining why they no longer view labor markets as inherently competitive (Kades 2023; Kanter 2023).

Search-and-Matching Models

For many decades, search-and-matching models were the primary lens through which labor economists understood labor market power. In a perfectly competitive market, workers would be paid a wage equal to their marginal revenue product of labor (often called the marginal product of labor for short). Search-and-matching models take seriously frictions in the labor market. In a frictional labor market, even if there are infinitely many employers, wages will be marked down relative to the marginal product of labor. The literature proposes many sources of the wage markdown, but they share a core idea: workers cannot simultaneously and costlessly solicit job offers from all employers. As a result, an employer can pay a worker a wage slightly below marginal product, because the worker does not always have access to a competing job offer that would cause that worker to leave (Burdett and Mortensen 1998) or force the incumbent employer to bid up the existing wage (Postel-Vinay and Robin 2002).

There is a rich literature on search-and-matching models; for excellent reviews, see Mortensen and Pissarides (1999) and Kline (2025). The literature convincingly documents the existence of labor market power arising from search-and-matching frictions. For the purposes of this article, however, this source of labor market power is of secondary interest. Antitrust enforcement cannot act directly on natural labor market frictions, unless they are amplified by specific prohibited conduct such as agreements between employers to avoid bidding wars.

Industrial Organization Models

Evidence from industrial organization models bears more directly on antitrust questions. The most relevant subset is a class of models that relates prices and quantities—or, in the case of labor markets, wages and employment—to the concentration of (noninfinitesimal) firms within a market. In a product market context, these models are familiar from undergraduate economics. As is well known, a monopolist seller of a good need not price at marginal cost in order to prevent consumers from switching to other sellers. The monopolist will therefore sell at a price above the competitive level, which also entails selling fewer units than the competitive quantity because demand is downward-sloping. Robinson (1933) was the first to apply this principle to labor markets, popularizing the term "monopsonist" to refer to a single employer that controls all the jobs in a given market (Thornton 2004). Like a monopolist, a labor monopsonist will set quantity (employment) below the competitive level. Because labor supply curves in this context are upward-sloping—offering higher (lower) wages will attract more (fewer) workers—this also entails wages below the competitive level. Perfect competition results in wages being bid up to the marginal product of labor. No employer will pay above the marginal product of labor because that would involve losing money on each worker, but no employer will pay below the marginal product of labor because then workers would switch to a competing employer that does pay the marginal product of labor.

A monopsonist, however, will find it profitable to pay a wage strictly below the marginal product of labor. In this model, if the firm raises the wage in order to attract an additional worker, it will also have to raise the wages of all its incumbent workers, increasing the effective cost of hiring the additional worker beyond the higher wage paid to that specific worker. The firm will internalize the fact that hiring an additional worker will raise the prevailing market wage, and therefore strategically set employment below the competitive level—or equivalently, set a wage below the marginal product of labor.

In practice, relatively few labor markets have just a single employer. However, many labor markets are concentrated, that is, dominated by a few large employers. The basic insights of the Robinsonian monopsony model extend to these more realistic cases. In a concentrated labor market, with a finite number of noninfinitesimal employers, each sizable employer affects the prevailing market wage. The firms will again internalize their impact on the prevailing wage, so they will strategically hire fewer workers than the competitive level, pushing wages below the marginal product of labor. Industrial organization models derive a tight relationship between market concentration and prices (for product markets) or wages (for labor markets). Under Cournot competition, the wage markdown (the marginal product of labor minus wages, divided by wages) is proportional to firm concentration.[5] Concentration is measured by the Herfindahl-Hirschman Index (HHI), defined as the sum

[5] Cournot competition refers to cases where firms compete by setting quantities. Firms can also compete by setting prices (Bertrand competition) or in more complex ways, but it is Cournot competition that generates a well-motivated theoretical link between concentration and wages.

of the squared market shares of all firms in a market. In this literature, a market is typically defined as combining a local area and a specific industry or occupation. Concentration-based models predict that concentration will reduce wages even in the absence of search-and-matching frictions.

In addition to concentration-based models of market power, another class of industrial organization models shows how market power can arise from firms producing differentiated products, which consumers view as imperfect substitutes. Consumers are then willing to pay a slightly higher price to buy a particular product whose characteristics they value more. A nascent empirical literature finds evidence that jobs, too, have an array of differentiated characteristics, so that employers can hire workers who are willing to accept a slightly lower wage for a particular job whose characteristics they value more. As a result, firms are able to pay workers a wage below the marginal product of labor (Azar, Berry, and Marinescu 2022; Lamadon, Mogstad, and Setzler 2022; Roussille and Scuderi 2025). Although the reality of differentiated jobs is not susceptible to policy intervention, this argument does nonetheless have implications for merger enforcement. If two merging firms that offer observably similar jobs in the same geographic area are not perfect substitutes from a worker's perspective, then a merger between these firms may not be as harmful to labor market competition as initially supposed. On the other hand, the definition of a labor market relevant to a worker may be narrower than the set of firms that offer observably similar jobs, in which case a merger may be more harmful to labor market competition.

Measuring Actionable Labor Market Power

Within the last decade or so, labor market power has begun to be understood more expansively than search-and-matching models alone (Card 2022). A series of recent papers document that many labor markets are concentrated. For example, Azar et al. (2020) and Azar, Marinescu, and Steinbaum (2022) construct measures of labor market concentration using shares of job postings as the definition of an employer's market share. They find that the average labor market is highly concentrated according to thresholds of the Herfindahl-Hirschman Index typically used in antitrust applications. For example, the average labor market HHI is 3,200 in CareerBuilder vacancy data (Azar, Marinescu, and Steinbaum 2022) or 4,400 in BurningGlass vacancy data (Azar et al. 2020). Under the 2023 Merger Guidelines, markets with an HHI above 1,800 are considered highly concentrated; this threshold was 2,500 under the 2010 Horizontal Merger Guidelines. Because large labor markets tend to be less concentrated, the average *worker* does not work in a highly concentrated labor market, defined as having an HHI above 2,500. The fraction of markets above this threshold is 60 percent, accounting for 16 percent of workers (Azar et al. 2020). These concentration calculations also use a labor market definition with a narrow definition of occupation (based on six-digit Standard Occupation Classification, or SOC, codes) that may raise the concentration beyond what

would be experienced by a worker who could switch to a closely related occupation. These six-digit SOC codes are fairly granular: there are more than 850 of them. A countervailing argument is that commuting zones, the geographic component of the local market definition used in these (and many other) papers, often cover multiple counties and in that sense are likely too broad for many workers. Nevertheless, these important papers challenge the long-held view that labor market concentration is negligible.

These authors and others then go a step further, studying the relationship between concentration and wages. An immediate challenge here is to find an exogenous source of variation for concentration. Azar, Marinescu, and Steinbaum (2022) use an instrumental variable approach, based on the number of posting employers in other geographic markets for the same occupation in a given quarter, which is intended to capture variation in market concentration driven by national-level rather than local factors. They find that going from the 25th to the 75th percentile of concentration within an occupation reduces posted wages by 17 percent.

This negative relationship between concentration and wages persists when using measures of actual worker pay rather than (incomplete) measures of wages from job postings. Other studies with related approaches and different data have found similar results, while offering additional insights. The concentration measures of these other studies map more cleanly to traditional product market concentration measures, because they use realized employment rather than job postings to calculate employer market shares. For example, Rinz (2022) uses actual earnings from individual tax filing data. Rinz's labor market definition uses industry (four-digit NAICS code) rather than occupation, and using the Longitudinal Business Database from 1976 to 2015 allows the author to document trends in within-industry labor market concentration over several decades. Although labor market concentration has been rising since the early 1990s when measured nationally, local concentration has been steadily falling since at least the mid-1970s, because much of the national increase is explained by large firms expanding across space. Again using an instrumental variable approach based on national-level variation, he finds that going from the 50th to the 75th percentile of concentration reduces earnings by about 10 percent. In another example, Qiu and Sojourner (2023) use earnings and health insurance benefits measured from the American Communities Survey and are able to control for product market concentration using data from Dun & Bradstreet. Controlling for product market concentration helps with the concern that large firms may increase both labor and product market concentration, with product market concentration potentially leading to higher product prices that can be passed through to higher wages. Using an instrumental variable approach similar to Azar, Marinescu, and Steinbaum (2022) and Rinz (2022), they find that moving up one standard deviation in labor market concentration lowers wages by 14 percent.

The constellation of findings from these early papers shows that there is a robust negative relationship between labor market concentration and wages. However, to determine whether this observed negative relationship offers a rationale for antitrust enforcement in labor markets, we need to reckon with the fact that concentration

and wages are both equilibrium outcomes (Bresnahan 1989; Miller et al. 2022). Both are determined by other underlying parameters, such as firm productivity. If the smallest firm in a market suddenly becomes more productive, then it will be able to pay higher wages and its market share will increase due to attracting more workers, which will reduce concentration. If, instead, the largest firm becomes more productive, then its market share will increase further, which will increase concentration while also increasing wages. In these examples, wages move in the same direction while concentration moves in opposite directions. The more fundamental point, however, is that neither one causes the other; they are both ultimately caused by changes in the firms' demand for labor. Thus, an observed negative relationship between concentration and wages may not be sufficient to conclude that reducing concentration truly causes wages to rise. This fundamental ambiguity remains even if one instruments for concentration (Miller et al. 2022), as the pioneering papers in the preceding paragraphs do.

The goal is then to isolate changes in the number or size of employers that change concentration but do not directly affect wages, except through their effect on labor market power. Some recent papers use mergers between employers in the same market to provide supporting evidence that higher employer concentration, rather than merely being correlated with lower wages, causes them. Prager and Schmitt (2021) study the effects of employer mergers on wages using a decade of hospital mergers. The authors find that the mergers that substantially increase labor market concentration cause subsequent annual wage growth to slow by one to two percentage points, or approximately one-quarter of baseline wage growth rates. However, they find no effects of mergers that only moderately increase concentration and no effects on occupations that are not specific to the hospital industry. Of course, focusing on the hospital industry limits generalizability. Using Census data to study employer mergers across the economy, Arnold (2025) finds that mergers that substantially increase labor market concentration cause earnings reductions of more than 2 percent. One complexity in merger-based studies is that a merger may also increase product market power, which allows firms to raise prices, thereby raising the marginal revenue product of labor. But when Arnold (2025) carries out separate estimates for tradeable industries, where product market power is unlikely to be meaningfully affected by a local merger, the negative effect of mergers on wages persists.

Additional papers that leverage mergers include Benmelech, Bergman, Kim (2022) and Guanziroli (2022), which similarly find slowdowns in wages, although Arnold et al. (2023) find declines only among workers at the acquired employer. The literature on how mergers affect labor markets is still young. Many important questions remain largely unanswered, such as distributional consequences, the interaction with product market power, and the role of organized labor.

A variety of other methods have been used to study the relationship between antitrust-actionable labor market power and wages. Several authors have built models that combine concentration and search-and-matching frameworks. When these authors bring their models to data, they find that concentration reduces wages. For example, Jarosch, Nimczik, and Sorkin (2024) develop and calibrate a model that

marries concentration and search frictions. They show that if employers can commit not to make new offers to the same worker in the future, then higher concentration results in lower wages. In another example, Berger, Herkenhoff, and Mongey (2022) marry concentration and job differentiation. Their model uses a representative household in a Cournot competition framework to show that both job differentiation and concentration result in lower wages. In both papers, the relevant measure of concentration is closely related to the standard Herfindahl-Hirschman Index measure.

Other studies turn to international and historical data. Delabastita and Rubens (2025) use a productivity estimation framework to estimate that collusion among nineteenth-century Belgian coal mines reduced wages by 6 to 17 percent. Sharma (2025) uses structural conduct testing methods to estimate that collusion among Indian textile firms suppresses wages by 10 percent. Gibson (2024) uses reduced-form methods to estimate that collusion among US tech companies in the mid-2000s reduced wages by 6 percent. Several studies show that noncompete agreements, which effectively increase labor market concentration by tying a worker to their current employer, seem to reduce wages.[6] The evidence on noncompete agreements is discussed in more detail in the paper by Starr in this symposium.

Taken together, these results provide strong evidence for the existence of labor market power that is within the purview of antitrust enforcers. Many sources of labor market power—high levels of concentration, search frictions, differentiated jobs—are outside the jurisdiction of antitrust law unless they are combined with specific violations. However, mergers, collusion, and other agreements to restrain competition are squarely within its jurisdiction.

Labor Market Considerations in Merger Enforcement

This article discusses two natural areas for antitrust enforcement in labor markets. This section discusses proposed mergers that might affect labor markets, and the next section discusses collusive agreements among employers to hold down wages or working conditions.[7] In each of these areas, there are unresolved issues that affect the practical implementation of enforcement.

Some have argued against considering labor market effects in merger review at all. The rationale is that protecting labor market competition will raise labor

[6] Noncompete agreements are distinct from the no-poaching agreements studied by Gibson (2024). Noncompete agreements are between an employer and a worker and, at least in principle, observable to the worker. No-poaching agreements are between employers and typically kept secret from workers.

[7] A third natural area is noncompete agreements between employers and workers, mentioned in the previous footnote. Another possible area is (the labor market analog of) monopolization, the gaining or maintenance of monopoly power through means other than "natural" sources of competitive advantage. An example might be predatory wage-setting: hiring away a competitor's key workers at a wage above the equilibrium wage the firm would pay if not for its attempt to gain market power by driving the competitor out of business. Little is known about the prevalence or effects of such conduct in labor markets, so it is omitted from this article.

costs, which will be passed through to higher prices and harm consumers. It is not obvious, however, that a merger that reduces labor market competition will generally benefit consumers. A key mechanism by which merging firms achieve wage reductions is to restrict the quantity of labor (Robinson 1933). Maintaining wages below the competitive level requires employing fewer workers, because the supply of workers is upward-sloping. Employing fewer workers also typically means producing less output, which then puts upward pressure on the final prices paid by consumers. Furthermore, labor market competition is important for allocative efficiency. Suppose, for example, that decades of hospital mergers weaken competition in the nursing labor market. Then in the long run, inefficiently few people may enter the nursing profession, because nurse compensation will be below the marginal product of nurse labor—which will likely make consumers worse off.

The exercise of labor market power could also manifest itself in lower wages, without a corresponding reduction in employment or output. A firm exercising labor market power will need employment reductions as a mechanism for lower wages only in markets where workers doing the same job are paid similarly to one another. Even in circumstances where pay is negotiated individually, however, the allocative efficiency argument still applies. Of course, long-run allocative efficiency effects are difficult for courts to evaluate. Nonetheless, they provide a principled economic argument in favor of applying merger enforcement tools to labor markets.

Antitrust enforcers have a set of standard tools to evaluate how a merger is likely to affect downstream purchasers, whether consumers or other firms. These tools can also be extended to how a merger might affect the market for inputs in general, and the market for labor inputs in particular. However, this last extension is not always straightforward. This section discusses the outstanding implementation questions that I view as most pressing or most difficult: in particular, issues related to market definition, what level of concentration should trigger scrutiny, and how to model job choice.

Market Definition

Plausible market definitions can be narrow or broad. A narrower market definition, such as Pilates studios in downtown Boston, can more easily result in the merging parties being seen as competing in a concentrated market. A broader market definition, such as all gyms and exercise classes in the Boston metropolitan area, results in the market being seen as less concentrated and therefore less likely to be harmed by a merger. To challenge a merger successfully, the antitrust agencies must typically show that the merger will harm competition in at least one "relevant" market. Mergers are therefore more likely to be allowed when the court accepts a broad market definition; indeed, the way in which a market is defined can, in effect, settle the question of whether a merger will raise anticompetitive concerns.[8]

[8]There are principled arguments against the entire exercise of market definition (Kaplow 2025), but due to the ongoing centrality of market definition to merger enforcement, this section takes as given the need for adapting it to labor markets.

In output market settings, the "hypothetical monopolist" test is often used to discipline which candidate market definitions can be considered relevant. This test deems a market definition valid if a hypothetical monopolist seller controlling the entire market would find it profitable to increase prices by some amount (typically 5 percent). The test is also sometimes referred to as the SSNIP test, because it relies on the calculation of a "small but significant and non-transitory increase in price." This test can be adapted for a labor market setting into the hypothetical monopsonist test (Naidu, Posner, and Weyl 2018). The test then asks whether a monopsonist employer would find it profitable to reduce wages by some amount. However, implementation of the test in a labor market setting is not straightforward. It requires estimation of the cross-employer wage elasticity of labor supply, which is complicated by the two-sided nature of labor markets (as discussed below in the section on models of job choice). In practice, the choice of market definition will often depend on the specifics of each case.

It is not always possible or desirable to invest in a bespoke labor market definition. At the initial stage of deciding whether to investigate a proposed merger, the antitrust agencies need an easily implementable labor market definition. If the merger is challenged in court, it is often desirable for the market definition to consist of standard components that are comprehensible to the judge and replicable by the opposing party. Table 1 summarizes some natural candidate definitions that are useful points of departure, although the most suitable definition may well vary on a case-by-case basis.

The choice of a definition necessitates trading off flexibility and accuracy against feasibility. Most of the candidate definitions in Table 1 rely on standard classification systems for industries, occupations, and geographies. Some of these classification systems, such as North American Industry Classification System (NAICS) codes and Standard Occupational Classification (SOC) codes, allow a choice of the level of granularity, because these codes are defined as broad categories divided into several levels of greater detail.

Each of these classification systems is subject to the criticism that it is not possible to partition labor markets fully: people do similar jobs across industries, switch occupations, and migrate across geographies. It may therefore be desirable to aggregate some cells into a coarser partition.[9] Consider the occupation dimension of labor market definition. One method for aggregation is to aggregate from the most granular six-digit SOC codes to broader five-, three-, or two-digit SOC codes. However, the degree of across-occupation switching is highly heterogeneous, and this is not always reflected in the nested structure of SOC codes.

An alternative method for aggregation is to construct a measure of occupation closeness. Figure 1 displays a subset of the results of a hierarchical clustering algorithm that groups occupations based on the frequency of occupation-to-occupation switches. The farther to the right two branches split, the more switching there is

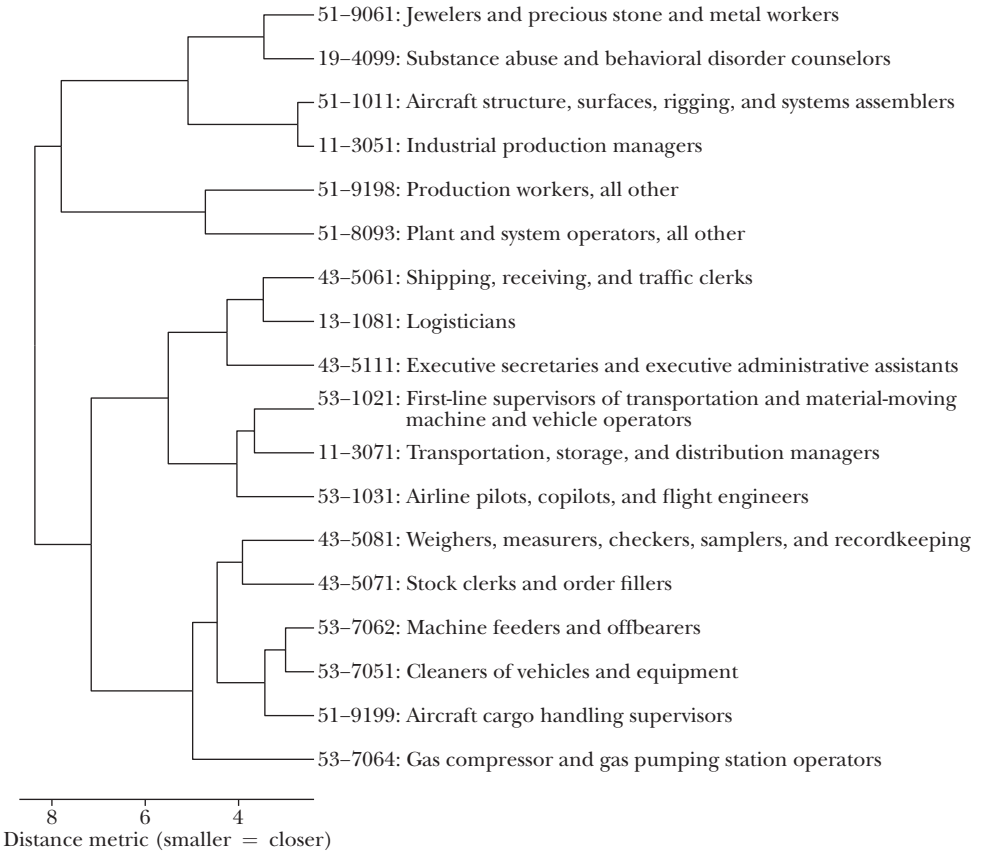
[9]Aggregations with a continuous component, as in Arnold (2025) or Jarosch, Nimczik, and Sorkin (2024), may have difficulty gaining acceptance by a US court under current precedent.

Table 1
Candidates for Defining a Labor Market

Basis for definition	Notes	Selected citations
Industry × geography	*Likely cell definition:* NAICS code × commuting zone. *Benefits:* Parallels market definition in early stages of product market merger evaluation. Feasible to compute in national datasets, because many include firm industry (for example, LBD, LEHD). *Drawbacks:* Ignores occupational boundaries. An accountant for a car manufacturing plant can more easily switch to an accounting job in another industry than to a factory floor job in car manufacturing. Commuting zones may not be the most relevant geographic delineations for workers (also applies to other candidate definitions in this table).	Arnold (2025); Berger, Herkenhoff, and Mongey (2022); Berger et al. (2025)
Occupation × geography	*Likely cell definition:* SOC code × commuting zone. *Benefits:* Occupation is likely more salient than industry from a worker's perspective, especially in licensed occupations. *Drawbacks:* Occupation is not available in US administrative data, so only feasible in other jurisdictions or using nonadministrative data (for example, resume data). Workers switch across occupations; this is partially addressable using cross-occupation flows (see main text).	Azar et al. (2020); Azar, Marinescu, and Steinbaum (2022); Qiu and Sojourner (2023)
Industry × occupation × geography	*Likely cell definition:* NAICS code × SOC code × commuting zone. *Benefits:* More flexible than preceding definitions, especially with appropriate aggregation across leaky industry × occupation cells as in Jarosch, Nimczik, and Sorkin (2024). *Drawbacks:* Occupation is not available in US administrative data and industry is often unavailable in data sources that contain occupation. Handwerker and Dey (2024) get as close as possible using internal Department of Labor data merged to other sources, but even those data are aggregated.	Handwerker and Dey (2024); Jarosch, Nimczik, and Sorkin (2024); Prager and Schmitt (2021)
Firm-to-firm flows	*Likely cell definition:* N/A. *Benefits:* Flexibility. For example, Herrera-Caicedo, Jeffers, and Prager (2025) propose a measure of labor market competition at the granular level of firm pairs that is continuous and allows heterogeneity by occupation. *Drawbacks:* Antitrust enforcers do not have access to firm-specific Census data for constructing such measures (for example, LEHD). If using resume data, then additional data cleaning burden from use of free-text fields (for example, firm name). Resume data often do not capture low-wage workers.	Herrera-Caicedo, Jeffers, and Prager (2025)

between the occupations in those branches. Some occupations with similar SOC codes, such as machine feeders and offbearers (53-7062) and cleaners of vehicles and equipment (53-7051), do indeed have very high occupation-to-occupation switching rates, as indicated by the short horizontal distance of the final branches that separate them. However, some occupations are very closely related despite being in different two-digit SOC codes, the highest level of SOC code aggregation. For example, shipping, receiving, and traffic clerks (43-5061) are closely related to logisticians (13-1081). Aggregation approaches such as this one provide a principled way to group together close SOC codes while leaving isolated codes in their own single-code groups (these are omitted from the figure for brevity). They can be combined with any of the candidate definitions in Table 1 that have an occupation component. This and other types of aggregation can be applied to other

Figure 1
Clustering SOC Occupation Codes by Cross-Occupation Flows



Source: Author's calculations based on occupation transition frequencies constructed by Schubert, Stansbury, and Taska (2021) from Burning Glass resume data.
Note: Branches that diverge at large *x*-axis values (near the left margin) indicate low occupation-to-occupation transition rates. The distance metric indicated on the *x*-axis is a monotonic transformation of the inverse of the pairwise occupation-to-occupation transition rate, calculated as a weighted average of the direction-specific transition rate. A distance of 8 maps to approximately a 0.03 percent occupation-to-occupation transition rate, and a distance of 2.5 maps to approximately an 8 percent transition rate. SOC codes 31-1011 and 11-3051, the closest two in this plot, split at a distance of approximately 2.7. Only branches from one selected cluster are shown. Clustering is implemented using average-linkage hierarchical clustering. Hierarchical clustering is sensitive to the choice of algorithm (single-linkage, complete-linkage, and others) and the distance metric.

dimensions of labor market definition besides occupation. For example, a similar idea is used by Herrera-Caicedo, Jeffers, and Prager (2025) to provide a continuous measure of how closely a pair of firms compete in the labor market.

Conversely, standard classifications are sometimes too broad. For example, French professors and Hindi professors, who likely compete for different jobs, are pooled into a single SOC code (25-1124), even when using the most granular (six-digit) SOC classification. The most commonly used geographic delineation, the commuting zones constructed by the US Department of Agriculture to capture cross-county commuting patterns, lacks a nested structure altogether, thus making it impossible to consider whether it is under- or over-inclusive. One end of a large urban commuting zone rarely sends commuters to the other end, even if both ends send commuters into the same central business district. Some workers living near borders may partially "belong" to more than one commuting zone. Moreover, workers migrate, and willingness to migrate is highly heterogeneous by age and education (Jia et al. 2023), and likely by occupation. On balance, the evidence that jobs are meaningfully differentiated even within industry-geography or occupation-geography cells (Azar, Berry, and Marinescu 2022; Lamadon, Mogstad, and Setzler 2022; Roussille and Scuderi 2025) points to appropriate labor markets being narrower than is typically assumed.

Labor Market Concentration Thresholds

The likelihood of a merger being challenged by US antitrust agencies increases substantially when the merger-induced increase in product market concentration and the post-merger level of concentration exceed thresholds set out in the Merger Guidelines. Under the most recent 2023 Guidelines, the thresholds for closer inspection of a proposed merger are a post-merger Herfindahl-Hirschman Index level above 1,800 and an HHI increase above 100.[10] While no single threshold is likely to be policy-optimal across all industries and markets, there is some support for the use of these thresholds for evaluating product market effects of mergers (Nocke and Whinston 2022).[11]

When it comes to labor markets, however, there is almost no evidence about the appropriate thresholds for concentration levels or increases. One possibility is that labor market harms begin at lower concentrations than product market harms, because labor markets may be characterized by greater market power. For example, information frictions are more severe than in most product markets. Unlike a product market where all prices are observable to consumers, labor markets are characterized by difficult-to-observe wages and working conditions. Learning these job attributes often requires not only completing a costly interview process but also securing a job offer. As noted earlier, these search frictions generate employer market power.

[10] This is a change from the 2010 Horizontal Merger Guidelines, whose thresholds were a post-merger HHI above 2,500 and an HHI increase above 200. The 2023 Guidelines revision returns to the 1982 Guidelines thresholds of 1,800 and 100.

[11] Even in product markets, the optimal thresholds are sensitive to assumptions, and the very concept of an HHI threshold is subject to the same critiques as market definition (Kaplow 2024). Given that the use of HHI thresholds is likely to persist for the foreseeable future, however, this article considers how best to implement them in labor markets.

In the study of local hospital mergers mentioned earlier, Prager and Schmitt (2021) find harms to labor market competition at Herfindahl-Hirschman Index increases of several hundred using a labor market definition with a narrow industry (hospitals) or approximately 100 using a broader industry (all health care). These estimates may be specific to the health care sector, which has low across-occupation mobility. Calibrating a model to many industries, Berger et al. (2025) provide suggestive evidence that the existing product market thresholds are in the right ballpark for labor markets.

Much more research is needed to determine the appropriate Herfindahl-Hirschman Index thresholds for evaluating labor market effects. The appropriate threshold will also be a function of the chosen labor market definition: a narrower market definition will produce a higher optimal threshold. Research that simultaneously considers labor market definition and HHI thresholds would therefore be especially valuable.

Alternatively, the constraints of legal precedent may mean that the same thresholds for concentration must be applied to both product and labor markets. The courts may be reluctant to recognize a separate set of thresholds for labor markets. One possible solution to this practical problem is to choose labor market definitions that are internally consistent with existing HHI thresholds, rather than to develop new thresholds.

Models of Job Choice

The final stages of a merger review can involve estimating a model of the markets in question in order to simulate the likely effects on prices, quantities, and sometimes other outcomes. Many merger cases, even many of those that eventually make it to court, do not involve such merger simulation. Merger simulation is costly and can be unnecessary in cases where other types of evidence clearly point to the merger eliminating or reducing strong competition. It is also sometimes infeasible given available data. Nevertheless, simulation of the relevant market remains one important tool for merger evaluation.

When merger simulation is undertaken, it is closely connected to economic theory and often resembles well-executed empirical industrial organization research. In standard product market merger cases, discrete choice models of product demand are often used. The direct labor market analog is to estimate a discrete choice model of workers' demand for jobs, likely allowing for some job differentiation, as in Azar, Berry, and Marinescu (2022). A worker is then modeled as choosing among a list of jobs with known wages and nonwage characteristics.

But even at a basic conceptual level, there are at least two difficulties with this analog. First, labor markets are matching markets. Unlike consumers choosing which of the available products to purchase, workers cannot choose among all existing jobs, because not every employer wants to hire a given worker. In typical datasets, it is difficult to infer for which jobs a worker would have been eligible. If the set of attainable jobs is defined too broadly, estimates of the wage elasticity of labor supply will be biased downward, yielding incorrect predictions about harms from

the merger. New models of job choice that combine the discrete choice approach with the matching nature of labor markets would be valuable for antitrust enforcers.

A second difficulty is that the relevant characteristics of a job are not easily observable to the analyst, and some are not even observable to the worker. Unlike in most product markets, the labor market analog of prices—that is, wages—is typically somewhat hidden from workers until after they have applied for and been offered a job. Given that workers do not know with certainty a key characteristic of most of the jobs that exist, it is difficult to estimate how they will respond when a characteristic changes due to a merger. Even if job characteristics were more fully known to workers, important dimensions such as the corporate culture, work hours, and schedule flexibility are often difficult for the analyst to measure. These information frictions complicate the estimation of discrete choice models. The industrial organization literature has made some progress toward incorporating such complications in product market settings, such as the literature on "consideration sets" that are used to study consumer choices between goods (Honka, Hortaçsu, and Wildenbeest 2019), but these tools have not yet been appropriately extended to labor market contexts.

Additional Considerations

Of the many other outstanding questions related to merger review of labor market power, I will mention three. First, there is evidence that labor unions mitigate the negative wage effects of employer power (Prager and Schmitt 2021; Benmelech, Bergman, and Kim 2022) and increase pass-through of employer product market power to higher wages (Dodini, Salvanes, and Willén 2024; CMA Microeconomics Unit 2024). An open question is whether and how merger evaluation should differ as a function of countervailing worker power.

Second, there is increasing attention paid to nonprice, nonquantity market outcomes such as quality and innovation in the evaluation of product market effects of mergers. The labor market analog is nonwage job amenities (for a review of the relevant literature, see Mas 2025). How to incorporate nonwage job amenities into merger analysis is an open question.

Finally, adding a labor market component to merger review is costly. The resource constraints of antitrust agencies are already binding (Brot et al. 2024), so adding this new component would crowd out other enforcement activity. Whether this resource shift is efficient hinges on what fraction of mergers that are not already flagged on other grounds would, upon investigation, turn out to be concerning on labor market grounds. This fraction is not known, although a promising early step toward calibrating it through the lens of a model is Hosken, Larson-Koester, and Taragin (2023). Nor is there a data infrastructure for easily evaluating the trade-offs. As noted earlier, substantial proposed mergers must be reported to the antitrust agencies for an initial review under the Hart-Scott-Rodino (HSR) premerger notification requirement. However, the HSR form only asks firms for information that indicates whether the merging parties have a substantial competitive overlap in product markets (often measured by industry codes and key product information). A proposed 2024 revision

to the HSR form would have begun to collect labor-related information (FTC 2024), but that portion of the revision was scuttled during the political process.

Restrictive Agreements among Employers

Agreements among employers that suppress competition for workers can be collusive. Typical examples include wage-fixing agreements and no-poaching agreements. Enforcement against collusion relies less heavily on detailed economic modeling than does merger enforcement, because the laws prohibiting collusion make it *per se* illegal: that is, illegal regardless of the magnitude or measurability of the harms. The *per se* prohibition is grounded in a robust prediction from economics that collusion harms markets, but the very strength of that prediction—through the *per se* prohibition—obviates the need for detailed empirical analysis. As such, the open questions are not so much about how to enforce laws against labor market collusion as they are about resource allocation. If labor market collusion is pervasive and easily detected, and its harms are large, then using limited antitrust enforcement resources on it is justified. If it is rare, minimally harmful, or very costly to detect, then limited resources should perhaps be targeted elsewhere.

An important caveat is that even these types of agreements may not constitute illegal collusion if they are "ancillary" to competition—that is, required for a procompetitive business relationship. For example, two firms engaging in a procompetitive joint venture may argue that without an agreement not to poach one another's relevant workers, they would not enter the joint venture in the first place.

Prevalence and Harms

The empirical evidence on the prevalence of labor market collusion and any resulting economic harms is in its infancy. Collusion in any market is difficult to study, because firms have a strong incentive to hide actions that are likely to be found illegal. This same property also makes its prevalence difficult to estimate. Nonetheless, there are documented instances of employer collusion dating as far back as the mid-1800s, as in the Delabastita and Rubens (2025) study of employer collusion between 227 Belgian coal firms in 1845–1913, and all the way up to modern times. Some recent examples include studies of the no-poaching agreements among prominent Silicon Valley firms (Gibson 2024; Herrera-Caicedo, Jeffers, and Prager 2025) and how large employers in India's textile and clothing industry organized into industry associations to hold down wages (Sharma 2025). The few empirical papers that exist in this area find wage reductions in the neighborhood of 5 to 15 percent. Whether this range is representative remains an open question. With this dearth of evidence on employer collusion, the appropriate level of enforcement resources remains uncertain.

Detection

When enforcement against collusive employer agreements does occur, it does not follow the definition of "collusion" used in economics. Economists tend to focus

on whether collusive outcomes were reached, without regard to whether explicit communication was involved in reaching them. For example, consider two airlines that refrain from lowering prices on routes where they directly compete, even when it would be individually profitable for them to lower prices. For economists, the outcome would be regarded as collusive regardless of whether the collusion was explicit (involving direct communication) or tacit. In contrast, the legal standard requires explicit communication or agreements between the firms. Direct communication likely makes collusion both easier and more effective (Green, Marshall, and Marx 2014; Awaya and Krishna 2016; Asker and Nocke 2021), so perhaps the welfare losses from allowing tacit collusion are comparatively small. However, collusion without direct communication is still feasible, so this legal standard leaves open legal avenues for welfare-reducing collusion (Harrington 2012; Byrne and de Roos 2019; Chassang and Ortner 2023). For example, a firm can experiment with keeping prices high in the face of falling demand in order to test whether its rival will also refrain from dropping prices.

Given these differences in how collusion is defined, the burden of proof for prosecuting collusion can seem quite high from an economist's perspective. It is not sufficient (or even necessary) to show that firms achieved collusive prices or other collusive outcomes; a history of communication must be shown. This high burden of proof applies to collusion in labor markets, as well. Historically, when antitrust enforcement in labor markets was weak, employers may have been rationally careless in failing to obfuscate their explicitly collusive behavior in labor markets. This may have given antitrust enforcers a temporary advantage in the labor market context, with employers leaving clear paper trails as described in Herrera-Caicedo, Jeffers, and Prager (2025).

It is reasonable to expect that the increasing focus of antitrust enforcement on labor markets will make firms both more reluctant to engage in labor market collusion and more motivated to hide it. An important question for researchers is therefore how to test for labor market collusion in the absence of a paper trail. In a product market context, many of the existing detection tools rely on suspicious pricing patterns, especially suspiciously stable prices (Harrington 2008). In labor markets, such tests may be made less useful by the fact that wage stability may naturally result from noncollusive sources such as downward wage stickiness, collective bargaining, and labor laws. Even after decades of research, detection of collusion in product markets remains challenging (Asker and Nocke 2021). Enforcers may have no choice but to rely more heavily on existing leniency programs. These programs have been successful in encouraging reporting by promising reduced penalties for the first firm that blows a whistle on collusion in which it is a participant (OECD 2003). In order for whistleblowing to be an attractive option, the penalties ordinarily imposed must be large. Under US law, fines for federally prosecuted collusion cannot exceed two times the firms' gain from the collusion (Hammond 2010). Even under these comparatively small fines, whistleblower programs have been effective (Miller 2009), but some argue that the fines should be larger (OECD 2003).

Conclusion

A decade ago, antitrust cases with a primary focus on labor markets were vanishingly rare. Labor market considerations now play a growing—though still small—role in antitrust enforcement. A recent research literature shows that labor market competition is less robust, and also more amenable to antitrust tools, than previously thought. New empirical evidence and modeling approaches that make progress on these questions can have a real impact on policy.

■ *I thank Susan Athey, Ioana Marinescu, Jeanine Miklos-Thal, Eric Posner, Kevin Rinz, and the JEP editorial team for helpful feedback.*

References

Arnold, David. 2025. "Mergers and Acquisitions, Local Labor Market Concentration, and Worker Outcomes." Unpublished.

Arnold, David, Kevin S. Milligan, Terry Moon, and Amirhossein Tavakoli. 2023. "Job Transitions and Employee Earnings after Acquisitions: Linking Corporate and Worker Outcomes." NBER Working Paper 31866.

Asker, John, and Volker Nocke. 2021. "Collusion, Mergers, and Related Antitrust Issues." In *Handbook of Industrial Organization*, Vol. 5, edited by Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, 177–279. North-Holland.

Athey, Susan, Mark Chicu, Malika Krishna, and Ioana Marinescu. 2023. "The Year in Review: Economics at the Antitrust Division, 2022–2023." *Review of Industrial Organization* 63 (4): 525–44.

Athey, Susan, Alex Gross, Ioana Marinescu, and Jennifer Shanefelter. 2024. "The Year in Review: Economics at the Antitrust Division 2023–2024." *Review of Industrial Organization* 65 (4): 809–29.

Awaya, Yu, and Vijay Krishna. 2016. "On Communication and Collusion." *American Economic Review* 106 (2): 285–315.

Azar, José A., Steven T. Berry, and Ioana Marinescu. 2022. "Estimating Labor Market Power." NBER Working Paper 30365.

Azar, José, Ioana Marinescu, and Marshall Steinbaum. 2022. "Labor Market Concentration." *Journal of Human Resources* 57 (S): S167–99.

Azar, José, Ioana Marinescu, Marshall Steinbaum, and Bledi Taska. 2020. "Concentration in US Labor Markets: Evidence from Online Vacancy Data." *Labour Economics* 66: 101886.

Benmelech, Efraim, Nittai K. Bergman, and Hyunseob Kim. 2022. "Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?" *Journal of Human Resources* 57 (S): S200–50.

Berger, David, Thomas Hasenzagl, Kyle Herkenhoff, Simon Mongey, and Eric A. Posner. 2025. "Merger Guidelines for the Labor Market." *Journal of Monetary Economics* 153: 103785.

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2022. "Labor Market Power." *American Economic Review* 112 (4): 1147–93.

Bresnahan, Timothy F. 1989. "Empirical Studies of Industries with Market Power." In *Handbook of Industrial Organization*, Vol. 2, edited by Richard Schmalensee and Robert Willig, 1011–57. North-Holland.

Brot, Zarek, Zack Cooper, Stuart V. Craig, and Lev Klarnet. 2024. "Is There Too Little Antitrust Enforcement in the US Hospital Sector?" *American Economic Review: Insights* 6 (4): 526–42.

Burdett, Kenneth, and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review* 39 (2): pp. 257–73.

Byrne, David P., and Nicolas de Roos. 2019. "Learning to Coordinate: A Study in Retail Gasoline." *American Economic Review* 109 (2): 591–619.

Card, David. 2022. "Who Set Your Wage?" *American Economic Review* 112 (4): 1075–90.

Chassang, Sylvain, and Juan Ortner. 2023. "Regulating Collusion." *Annual Review of Economics* 15: 177–204.

CMA Microeconomics Unit. 2024. *Competition and Market Power in UK Labour Markets*. Competition and Markets Authority.

Delabastita, Vincent, and Michael Rubens. 2025. "Colluding against Workers." *Journal of Political Economy* 133 (6): 1796–839.

Dodini, Samuel, Kjell Salvanes, and Alexander Willén. 2024. "The Dynamics of Power in Labor Markets: Monopolistic Unions versus Monopsonistic Employers." Unpublished.

Dunne, Timothy, Mark J. Roberts, and Larry Samuelson. 1988. "Patterns of Firm Entry and Exit in US Manufacturing Industries." *RAND Journal of Economics* 19 (4): 495–515.

FTC. 2024. *16 CFR Parts 801 and 803. RIN 3084-AB46. Premerger Notification; Reporting and Waiting Period Requirements*. Federal Trade Commission, Billing Code 6750-01-P.

Gibson, Matthew D. 2024. "Employer Market Power in Silicon Valley." Upjohn Institute Working Paper 24–398.

Green, Edward J., Robert C. Marshall, and Leslie M. Marx. 2014. "Tacit Collusion in Oligopoly." In *The Oxford Handbook of International Antitrust Economics*, Vol. 2, edited by Roger D. Blair and D. Daniel Sokol, 464–97. Oxford University Press.

Guanziroli, Tomas. 2022. "Does Labor Market Concentration Decrease Wages? Evidence from a Retail Pharmacy Merger." Unpublished.

Hammond, Scott D. 2010. "The Evolution of Criminal Antitrust Enforcement Over the Last Two Decades." Speech, presented at the Twenty-Fourth Annual National Institute on White Collar Crime, Miami, FL, February 25, 2010. https://www.justice.gov/d9/atr/speeches/attachments/2015/06/25/the_evolution_of_criminal_antitrust_enforcement_over_the_last_two_decades.pdf.

Handwerker, Elizabeth Weber, and Matthew Dey. 2024. "Some Facts about Concentrated Labor Markets in the United States." *Industrial Relations: A Journal of Economy and Society* 63 (2): 132–51.

Harrington, Joseph E., Jr. 2008. "Detecting Cartels." In *Handbook of Antitrust Economics*, edited by Paolo Buccirossi, 213–58. MIT Press.

Harrington, Joseph. 2012. "Game-Theoretic Ruminations on Section 1 of the Sherman Act." Speech, Bayard Wickliffe Heath Memorial Lecture, University of Florida Levin College of Law, March 30, 2012.

Herrera-Caicedo, Alejandro, Jessica Jeffers, and Elena Prager. 2025. "Collusion through Common Leadership." NBER Working Paper 33866.

Honka, Elisabeth, Ali Hortaçsu, and Matthijs Wildenbeest. 2019. "Empirical Search and Consideration Sets." In *Handbook of the Economics of Marketing*, Vol. 1, edited by Jean-Pierre Dubé and Peter E. Rossi, 193–257. North-Holland.

Hosken, Daniel, Miriam Larson-Koester, and Charles Taragin. 2023. "Labor and Product Market Effects of Mergers." Unpublished.

***In re: High-Tech Employee Antitrust Litigation*.** US District Court for the Northern District of California, No. 5:11-cv-02509 (2011).

Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin. 2024. "Granular Search, Market Structure, and Wages." *Review of Economic Studies* 91 (6): 3569–607.

Jia, Ning, Raven Molloy, Christopher Smith, and Abigail Wozniak. 2023. "The Economics of Internal Migration: Advances and Policy Questions." *Journal of Economic Literature* 61 (1): 144–80.

Kades, Michael. 2023. "Antitrust Division Deputy Assistant Attorney General's Remarks." Speech, Annual CRA Brussels Conference, Brussels, December 7, 2023. https://www.justice.gov/archives/opa/speech/antitrust-division-deputy-assistant-attorney-general-michael-kades-gives-remarks-annual.

Kanter, Jonathan. 2023. "Assistant Attorney General's Remarks." Speech, Georgetown Antitrust Law Symposium, Washington, DC, September 19, 2023. https://www.justice.gov/archives/opa/speech/assistant-attorney-general-jonathan-kanter-delivers-remarks-2023-georgetown-antitrust.

Kaplow, Louis. 2024. *Rethinking Merger Analysis*. MIT Press.

Kaplow, Louis. 2025. "Improving Economic Analysis in Merger Guidelines." *Journal of Economic Perspectives* 39 (1): 29–52.

Kariel, Joel, Jakob Schneebacher, and Mike Walker. 2024. "Competition Policy and Labour Market Power: New Evidence and Open Questions." *Oxford Review of Economic Policy* 40 (4): 787–807.

Kline, Patrick M. 2025. "Labor Market Monopsony: Fundamentals and Frontiers." NBER Working Paper 33467.

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.

Mas, Alexandre. 2025. "Non-wage Amenities." NBER Working Paper 33643.

Miller, Nathan H. 2009. "Strategic Leniency and Cartel Enforcement." *American Economic Review* 99 (3): 750–68.

Miller, Nathan, Steven Berry, Fiona Scott Morton, Jonathan Baker, Timothy Bresnahan, Martin Gaynor, Richard Gilbert, et al. 2022. "On the Misuse of Regressions of Price on the HHI in Merger Review." *Journal of Antitrust Enforcement* 10 (2): 248–59.

Mortensen, Dale T., and Christopher A. Pissarides. 1999. "New Developments in Models of Search in the Labor Market." In *Handbook of Labor Economics*, Vol. 3, edited by Orley C. Ashenfelter and David Card, 2567–627. North-Holland.

Naidu, Suresh, Eric A. Posner, and Glen Weyl. 2018. "Antitrust Remedies for Labor Market Power." *Harvard Law Review* 132 (2): 536–601.

Nocke, Volker, and Michael D. Whinston. 2022. "Concentration Thresholds for Horizontal Mergers." *American Economic Review* 112 (6): 1915–48.

OECD. 2003. *Hard Core Cartels: Recent Progress and Challenges Ahead*. OECD.

Posner, Eric A. 2023. "The New Labor Antitrust." *Antitrust Law Journal* 86 (2): 503–78.

Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica* 70 (6): 2295–350.

Prager, Elena. 2026. *Data and Code for: "Antitrust Enforcement in Labor Markets."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E241085V1.

Prager, Elena, and Matt Schmitt. 2021. "Employer Consolidation and Wages: Evidence from Hospitals." *American Economic Review* 111 (2) 397–427.

Qiu, Yue, and Aaron Sojourner. 2023. "Labor-Market Concentration and Labor Compensation." *ILR Review* 76 (3): 475–503.

Rinz, Kevin. 2022. "Labor Market Concentration, Earnings, and Inequality." *Journal of Human Resources* 57 (S): S251–83.

Robinson, Joan. 1933. *The Economics of Imperfect Competition*. Macmillan.

Roussille, Nina, and Benjamin Scuderi. 2025. "Bidding for Talent: A Test of Conduct in a High-Wage Labor Market." NBER Working Paper 33848.

Schubert, Gregor, Anna Stansbury, and Bledi Taska. 2021. *Occupational mobility dataset for "Employer Concentration and Outside Options."* Distributed by: Author's website, https://annastansbury.github.io/website/Occ%20Transitions%20Public%20Data%20Set%20%28Jan%202021%29 (accessed June 25, 2025).

Sharma, Garima. 2025. "Collusion among Employers in India." Unpublished.

Starc, Amanda, and Thomas G. Wollmann. 2025. "Does Entry Remedy Collusion? Evidence from the Generic Prescription Drug Cartel." *American Economic Review* 115 (5): 1400–438.

Thornton, Robert J. 2004. "Retrospectives: How Joan Robinson and BL Hallward Named Monopsony." *Journal of Economic Perspectives* 18 (2): 257–61.

Thurman Arnold Project. 2020. "Modern Antitrust Enforcement." https://som.yale.edu/centers/thurman-arnold-project-at-yale/modern-antitrust-enforcement.

United States v. Activision Blizzard, 1:23-cv-00895 (2023).

United States v. Bertelsmann SE and CO KGaA, et al., 21-2886-FYP (2022).

United States v. Cargill, 1:22cv-01821 (D. Md. 2022).

United States v. DaVita Inc., 1:21-cr00229, (D. Colo. 2021).

United States v. Jindal, 4:20-cr-358 ALM/KPJ (2020).

United States v. Lopez, 22-40121 (5th Cir. 2023).

United States v. Manahe, 2:22-cr-00013, (D. Me. 2023).

United States v. Mahesh Patel, et al., 3:21-cr-220 (VAB) (2021).

United States v. VDA, OC, LLC, formerly Advantage on Call, LLC, 2:21-cr-00098-RFB-BNW (2021).

US Department of Justice, and Federal Trade Commission. 2023. *2023 Merger Guidelines*. US Department of Justice and the Federal Trade Commission.

Journal of Economic Perspectives—Volume 40, Number 1—Winter 2026—Pages 139–166

The Economics of Noncompete Clauses

Evan Starr

For centuries, policymakers and judges have debated the questions raised by noncompete clauses—contract terms that prohibit departing workers from joining or starting a competing firm, often within specified time and geographic limits (Blake 1960). The earliest reported case, *The Dyer's Case* of 1414, involved an apprentice named John Dyer who promised his master that he would not practice for six months in the same town. The master sued, claiming that the promise had been violated, but then failed to appear in court. Judge Hull held that common law did not recognize a right to restrain trade, and remarked that had the master been present to make such a claim, "he would go to prison until he paid a fine to the King" (Blake 1960, n. 33). By the early eighteenth century, however, English courts softened their stance. In *Mitchell v. Reynolds* (1711), Mitchell rented his bakeshop to Reynolds, with the provision that Reynolds would not act as a baker at another location in the parish for the next five years. However, Reynolds did start work at a different bakery, and Lord Macclesfield held that the contract provision was enforceable because Mitchell would not have been willing to rent the bakeshop unless assured that existing customers would be able to continue in their habit of purchasing baked goods at this location. Lord Macclesfield thus established a principle that still guides modern law: a covenant not to compete may be enforceable if it is "reasonable" in scope (FTC 2023a).

This "reasonableness" rule crossed the Atlantic and became the doctrinal foundation of US law. In most states, courts will uphold a noncompete if it protects a "legitimate business interest"—typically trade secrets or customer goodwill—without

■ *Evan Starr is a Professor of Management & Organization, Robert H. Smith School of Business, University of Maryland, College Park, Maryland. His email address is estarr@umd.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251457.

imposing undue harm on the worker or society. What counts as "reasonable," however, has varied widely across states (Bishara 2011). Some states enforce noncompetes even after workers are fired, while others do not. Some states allow judges to rewrite overly broad contracts to make them enforceable, while others strike down such contracts entirely. When a noncompete is signed during employment, some states require additional compensation to be paid to the employee for the new restriction, while others view continued employment as sufficient consideration. Notably, California, North Dakota, and Oklahoma prohibited employee noncompetes altogether in the late nineteenth century (Arnow-Richman 2019), while Minnesota banned them in 2023 (Narayan and Nunn 2025).

The ongoing debate over whether to allow noncompete clauses, or under what terms to do so, carries significant implications. Critics emphasize that noncompetes limit mobility, entrepreneurship, and innovation. Some liken them to violations of the Thirteenth Amendment (Hardaway 2016), while others attribute part of Silicon Valley's success to California's 1872 ban on such clauses (Gilson 1999; Saxenian 1996; Hyde 2003). Proponents, by contrast, argue that noncompetes represent voluntary contracting and can raise innovation and wages by encouraging firms to invest in development of trade secrets and training (Rubin and Shedd 1981).

This article draws on recent theory and evidence to revisit the central questions: Who signs noncompetes? Do workers benefit or suffer from them? Do firms truly need them, or can narrower tools suffice? Do noncompetes boost or stifle innovation? Do they help or harm small businesses? Should executives be included in bans on noncompete clauses? Do consumers ultimately benefit or lose? And is the status quo sufficient to mitigate existing harms from noncompetes? Following a brief overview of the relevant economic and legal background, I address these questions.[1] The weight of the evidence suggests that while narrow circumstances may exist in which noncompetes are privately efficient, their broader pattern is to function as restraints of trade in labor and product markets, suppressing wages, mobility, entrepreneurship, and innovation while raising costs to consumers.

Background

Relevant Economic Theories

Debates over noncompete clauses center on two competing, though not mutually exclusive, perspectives (Blake 1960). One sees noncompetes as efficient

[1] In different formats and forums, I have discussed this literature before, and this essay draws to some extent on these earlier writings. This includes written and oral testimony to the US Senate (Starr 2019a) and the US House of Representatives (Starr 2019b), in my comment to the FTC (Starr 2023b), a brief in 2019 for the Economic Innovation Group (Starr 2019c) (paid), an updated brief for the Economic Innovation Group in 2023 (Starr 2023a) (paid), an amicus brief (Starr 2024) submitted to the US Texas District Court in the *FTC v. Ryan* case and the *FTC v. Villages* case, and an amicus brief (Starr 2025) submitted to the appellate court in the *FTC v. Ryan* case. I also served as an observer on the Uniform Law Commission that developed the Uniform Restrictive Employment Agreement Act.

contracting devices; the other regards them as anticompetitive restraints that amplify the labor and product market power of firms.

The efficient contracting view, articulated by Rubin and Shedd (1981), holds that noncompetes can benefit both workers and firms by mitigating holdup problems (Grossman and Hart 1986). When a firm invests in developing valuable knowledge—such as the formula for Coca-Cola—that knowledge makes its employees more productive once shared. But because employees can carry the knowledge with them to a competitor, rival firms may benefit without paying the costs of innovation. Anticipating this leakage, the original firm may underinvest in training or research, resulting in lower productivity and lower wages. Noncompetes offer a blunt solution: by preventing workers from moving to competitors for a limited period, they help ensure the firm can capture the returns to its investment. In turn, the firm shares some of those returns with the worker to compensate them for the loss of their postemployment freedom.

The alternative view focuses on the anticompetitive nature of noncompetes. This view emphasizes that, just like search frictions create market imperfections that stem mobility and reduce wages relative to a competitive benchmark (Burdett and Mortensen 1998; Manning 2013), noncompetes exacerbate such imperfections by imposing competitor-specific moving costs. From this perspective, noncompetes are less about protecting investment and more about restricting turnover and curtailing wage costs. This perspective generally assumes that labor markets are sufficiently imperfect that workers have little choice but to accept noncompetes when presented (for example, when presented after they have already accepted a job). Once in place, a noncompete can insulate a firm from labor market competition for the duration of the employee's tenure until the expiration of the agreement, effectively extending any initial labor market monopsony power. In this way, noncompetes may misallocate labor, reduce wages, and reduce overall welfare by preventing workers from moving to jobs or starting firms where they are most productive.

Relevant Legal Background and History

The tension between these two economic perspectives has long been mirrored in a patchwork of state-level US laws. However, developments in the last decade or so have disrupted the status quo. In the mid-2010s, following the discovery of noncompetes in surprising places and early evidence on the use and effects of noncompetes, the US Department of the Treasury (2016) and White House (2016) issued policy reports urging reform and highlighting best practices for state law. Since then, hundreds of state bills have been introduced, with many passing (Bishara and Luisetto 2025). These state reforms included banning noncompetes for all or some workers (for example, for low-wage workers, physicians, or workers in high-technology industries), limiting employers from presenting noncompetes after employment started, and imposing penalties on firms that violated the law (for a review of states that changed their noncompete laws since 2011, see Fair Competition Law 2025).

In 2024, the Federal Trade Commission went further, proposing a nationwide rule that would prohibit nearly all noncompetes (FTC 2024). The agency

received more than 26,000 public comments on its proposed rule, as well as many amicus briefs (FTC 2023a). The US District Court for the Northern District of Texas ultimately vacated the proposed rule in *Ryan LLC v. Federal Trade Commission* (Docket Number 3:24-cv-00986-E, August 20, 2024) on the ground that a ban on noncompetes exceeded the statutory authority of the FTC. Thus, notwithstanding FTC enforcement actions against specific companies for noncompete clauses, state law continues to define the landscape of noncompete enforcement.

The documentary record from the proposed federal rule to ban noncompetes, as well as the public record from similar proposed noncompete bans at the state level, provide the clearest articulation of the crucial questions in the modern debate over noncompete clauses. In the following sections, I review the main questions and what recent literature has to say about the answers.

Who Signs Noncompetes?

In the mid-2010s, several news articles revealed the presence of noncompete clauses in seemingly unlikely jobs, including minimum-wage sandwich makers at Jimmy John's (as reported by Jamieson 2014), temporary packers at Amazon (reported by Woodman 2015), and summer camp counselors (reported by Greenhouse 2014). While these anecdotes drew attention to the spread of noncompetes, critics argued that such cases were "outliers." For example, in an amicus brief challenging the FTC's proposed rule, the National Retail Federation (Weibust et al. 2024) claimed that "[n]oncompetes are not regularly used for low wage workers," and that even in unusual cases "statutory and equitable protections [are] already in place for employees." Are noncompetes widespread or not? Who actually signs noncompetes?

Economic theory offers distinct predictions. The efficient contracting view suggests that noncompetes should be concentrated in jobs where employees have access to valuable and portable resources—trade secrets, confidential information, client lists, or specialized training (Callahan 1985). These are the types of "legitimate business interests" that courts traditionally recognize when enforcing a noncompete. Further, noncompetes should be found in places that would enforce such restrictions (and thus protect the underlying firm investments). By contrast, the anticompetitive view—that noncompetes are used to stifle turnover and wage growth—predicts much broader use: firms will impose noncompetes across many roles, including low-wage positions, at moments when workers have little bargaining power, and even where contracts are unlikely to survive judicial review.

The primary mechanism by which we have learned about the use of noncompetes are nationally-representative and occupation-specific surveys of employees and employers. Employer surveys suggest that many firms—though not all—use noncompetes indiscriminately. In a 2017 national survey of 634 private-sector firms with more than 50 employees, Colvin and Shierholz (2019) found that nearly one-third required noncompetes of all workers, while about half used them for some workers. These practices were not confined to high-paying firms: among employers

with average hourly wages below $13, approximately 30 percent required noncompetes for every worker. A separate 2017 national survey of roughly 1,500 firms found nearly identical results, with 29.5 percent requiring noncompetes for all employees and two-thirds for at least some (Balasubramanian, Starr, and Yamaguchi 2024). A Government Accountability Office (2023) survey of 446 firms in the Society for Human Resource Management database similarly found that 55 percent used noncompetes for some workers, and among those employing hourly staff, more than half covered all hourly workers with noncompetes. These boilerplate practices likely explain the cases of janitors (reported in Shubber 2018), unpaid interns (Lucas 2019), and even independent contractor window cleaners being asked to sign noncompetes (reported in Cowgill, Freiberg, and Starr 2024).

Employee surveys reinforce the picture that noncompetes are widely used. A nationally representative 2014 survey of 11,500 workers (Starr, Prescott, and Bishara 2021) found that 18 percent of workers overall were bound by a noncompete. Prevalence rose to about one-third among those with access to trade secrets, but even among those without such access, about 15 percent had signed one. Other nationally representative studies similarly show that noncompetes are more common among workers with higher wages, more education, and jobs in technical, managerial, or sales roles, yet they remain present among low-wage workers and those with little formal education (Starr, Prescott, and Bishara 2021; Krueger and Posner 2018; Balasubramanian, Starr, and Yamaguchi 2024; Boesch et al. 2023). For example, using the National Longitudinal Survey of Youth (NLSY), Rothstein and Starr (2022) find that 14 percent of workers below the median wage have a noncompete, including about 10 percent of the lowest wage earners. Half of all noncompete signers in the NLSY earned $23 per hour or less.

Occupation-specific studies confirm this duality: heavy use in high-skill jobs, but also notable presence in lower-wage occupations. Roughly 45 percent of physicians (Lavetti, Simon, and White 2020), 43 percent of engineers (Marx 2011), and 60 to 80 percent of executives (Schwab and Thomas 2006; Bishara, Martin, and Thomas 2015; Shi 2023) report having noncompetes. But about 30 percent of salon owners required their most recently hired hairstylist to sign one (Johnson and Lipsitz 2022), illustrating how such contracts reach well beyond elite professions.

Employers frequently deploy noncompetes even in states where courts will not enforce them. Anecdotal examples of noncompetes from California, where such contracts are banned, include a volunteer position at a nonprofit dedicated to young girls' fitness based in Silicon Valley and independent contractors at a moving company based in Huntington Beach (Cowgill, Freiberg, and Starr 2024). Indeed, nearly 30 percent of California workplaces required all workers to sign noncompetes, despite California's longstanding statutory prohibition (Colvin and Shierholz 2019). More broadly, prevalence of noncompetes is only modestly lower in states that prohibit enforcement (Rothstein and Starr 2022; Starr, Prescott, and Bishara 2021). Similarly, occupation-specific studies show slightly higher use of noncompetes in states that allow enforcement, but still find significant incidence in states that do not (Sanga 2018; Lavetti, Simon, and White 2020). As Fisk (2001)

notes, employers may be relying on the *in terrorem* effect: even if unenforceable, the presence of a contract may deter workers from seeking new jobs or convince them that they are legally bound.

Firms also frequently present noncompetes at inopportune times. Marx (2011) finds that about 70 percent of engineers in his sample were first asked to sign a noncompete after accepting a job offer, with nearly half on or after their first day of work. Similarly, Starr, Prescott, and Bishara (2021) report that roughly one in three workers signed under similar circumstances. Though such late notice may be anticipated in some cases, Marx concludes (p. 706), "the process by which firms obtain non-compete signatures suggests that firms do not use noncompetes merely for the oft-stated reason of protecting trade secrets and research and development investment."

Overall, while noncompetes are more common in jobs that involve trade secrets and high wages, their use is far broader than efficiency-based theories would predict. Many firms adopt them indiscriminately, impose them at inopportune times, and continue to use them even where courts would not uphold them. A recent review of noncompetes in Italy, the United Kingdom, Australia, and other OECD countries broadly suggests that these US-based findings are not anomalistic; rather, they reflect many workers experiences across the globe (Boeri et al. 2025; Majeed et al. 2025; CMA Microeconomics Unit 2024; Andrews and Garnero 2025).

Do Workers Benefit or Suffer from Noncompetes?

The efficient contracting and anticompetitive perspectives on noncompetes offer opposing predictions on whether they harm workers.

The anticompetitive view is often described as noncompetes trapping workers in jobs, preventing them from taking better jobs with higher earnings. As one example, a commenter to the Federal Trade Commission (2024) described her experience working as a bartender under a noncompete:

> In October 2020, I started working as a bartender at a company called [REDACTED] for $10 an hour. On my first day, I unknowingly signed a 2-year non-compete, slipped between other paperwork while my boss rushed me, and downplayed its importance. . . . At [REDACTED], I was sexually harassed and emotionally abused. I needed money, so I searched for a new job while remaining at [REDACTED] for one year. I was eventually offered a bartending job at a family-owned bar with better wages, conditions, and opportunities. Upon resigning, I was threatened with a non-compete I didn't know existed. Still, I couldn't take it anymore, so believing it was an unenforceable scare tactic, I took the new job, thinking our legal system wouldn't allow a massive company with over 20 locations to sue a young entry-level worker with no degree. In December 2021, I was sued for $30,000 in "considerable and irreparable damages" for violating the non-compete.

In contrast, the Retail Industry Leaders Association (2024) wrote in their comment opposing the proposed ban by the Federal Trade Commission that noncompetes (p. 4) "provide strong incentives for companies to invest in and financially reward employees, as they ensure to a reasonable degree that employers benefit from training employees and that confidential business information—including business strategies—and company trade secrets will not be taken to competitors immediately upon the departure of an employee." The National Retail Federation Amicus Brief (Weibust et al. 2024) simply claims: "Noncompetes do not reduce workers' wages." These claims mirror those in the academic literature espousing noncompetes as efficient contracting choices (Rubin and Shedd 1981; Meese 2022).

Complicating these competing predictions, a cursory review of the empirical evidence on how noncompetes relate to wages is indeed mixed (Starr 2021). As a result, some commentators suggest that the scientific evidence has been "muddled" (Torgnes and Terjesen 2025) and the policymakers have been accused of "cherry-picking" the evidence (for example, Wilson 2023). However, the "muddled" evidence appears to be the result of correlational studies, while studies plausibly estimating the causal effects of noncompetes (or noncompete enforceability) tend to point toward noncompetes reducing wages.

Correlational studies tend to find that workers with noncompetes earn more than workers without noncompetes (Lavetti, Simon, or White 2020; Starr, Prescott, and Bishara 2021; Rothstein and Starr 2022; Shi 2023; Gopal, Li, and Rawling 2025; Kodama, Kambayashi, and Izumi 2025). However, these correlations are difficult to interpret because firms typically adopt noncompetes alongside other restrictions like nondisclosure and nonsolicitation agreements (Balasubramanian, Starr, and Yamaguchi 2024; Boeri, Garnero, and Luisetto 2025; Majeed et al. 2025), and they tend to use such agreements more in higher-paying, technical jobs. This combination of factors makes it hard to know whether noncompetes themselves are responsible for higher wages. Using survey data collected by the compensation software company Payscale.com, Balasubramanian, Starr, and Yamaguchi (2024) show that workers with noncompetes almost always have other restrictions alongside them—again, nondisclosure agreements and nonsolicitation clauses—and that workers with all those restrictions earn 3–7 percent less than workers with only a nondisclosure agreement, even though they earn more than workers with no restrictions at all. This decomposition clarifies that (1) the apparent noncompete-wage premium reflects selection into any restrictions and (2) that observational comparisons of workers with and without a noncompete are confounded by (at least) the simultaneous adoption of other restrictions. Moreover, several studies find that noncompete-wage premia are actually larger in states where noncompetes are unenforceable (Starr, Prescott, and Bishara 2021; Rothstein and Starr 2022; Boeri, Garnero, and Luisetto 2025)—a pattern that does not fit the efficient contracting narrative.[2] Finally, workers report

[2]Two studies offer notable exceptions to this conclusion. Lavetti, Simon, and White (2020) perform a similar analysis with physicians and find an opposite effect (that the noncompete-wage premium is

rarely negotiating over the terms of their noncompetes or for benefits in exchange for signing (Starr, Prescott, and Bishara 2021) and wage bargaining is unrelated to noncompete use (Rothstein and Starr 2022; Cowgill, Freiberg, and Starr 2024).

In contrast to these mixed results from correlational studies, researchers who have used field experiments and natural experiments almost nearly uniformly find that noncompete clauses or noncompete enforceability suppress wages. The most common approach has been to use the state policy variation as natural experiments, with some studies focusing on smaller changes in noncompete laws (for example, Johnson, Lavetti, and Lipsitz 2025) and others focusing on the noncompete bans passed in the last two decades. In 2008, for example, Oregon banned noncompete clauses for low-wage workers (Lipsitz and Starr 2022); in 2015, Hawaii banned noncompetes (alongside agreements not to solicit coworkers) for high-tech workers (Balasubramanian et al. 2022); and in 2020, Washington banned noncompetes for workers earning under $100,000 (Hiraiwa, Lipsitz, and Starr 2024). Researchers examining all such bans (Greenwood, Kobayashi, and Starr forthcoming), as well as those studying a multitude of smaller changes to state noncompete laws (Johnson, Lavetti, and Lipsitz 2025), come to the same conclusion: banning noncompetes spurs mobility and increases average wages, typically by around 3–4 percent.[3]

These estimates likely underestimate the extent to which removing noncompetes from contracts will increase wages. Even when noncompetes are banned in a state, firms still use unenforceable noncompetes, and even unenforceable noncompetes still chill worker mobility (Starr, Prescott, and Bishara 2020; Cowgill, Freiberg, and Starr 2024). The use of unenforceable noncompetes also implies that studying state-level policy variation alone cannot estimate how actual noncompetes affect workers or firms.

Given that noncompetes often come bundled with other restrictions, Balasubramanian, Starr, and Yamaguchi (2024) suggest that estimating the causal effect of a firm requiring—or a worker signing—a noncompete may require a field experiment approach in which the terms of the contract can be specifically modified to remove or include *only* a noncompete. Thus, in Cowgill, Freiberg, and Starr (2024), the researchers made approximately 14,000 job offers on behalf of a firm ("Firm A") with short-term human resources contractors in the finance industry, randomizing the presence, salience, and duration of a noncompete in the job offer

higher in states that enforce noncompetes), though their analysis is constrained by having only five states in the data. Kini, Williams, and Yin (2021) perform a similar analysis for executives and find similarly that the noncompete-compensation differential is more positive in states that enforce noncompetes, but they find a strong negative relationship between noncompetes and compensation at the baseline.

[3] International evidence is slightly more mixed. Young (2024), studying a low-wage noncompete ban in Austria, finds that it spurs mobility but has no effect on wages. In contrast, Bartelsman, Dobbelaere, and Mattioli (2024) finds that the 2015 Dutch decision not to enforce noncompetes for temporary contracts increased labor income by 13 percent. Note also that in a recent working paper Narayan and Nunn (2025) find very short-run null effects of Minnesota's 2023 ban on noncompetes, though this may be expected as the law was not retroactive.

alongside the wage offer itself. When the noncompete is on page 7 of the contract, Cowgill, Freiberg, and Starr (2024) find that workers largely do not read it, with approximately one-third skipping it altogether, leading to no detectable differences in job completion rates for workers with versus without noncompetes in their job offer. After finishing the job with Firm A, the researchers find that those with noncompetes are 57 percent less likely to work for a competitor firm ("Firm B"). Altogether, the causal effect of Firm A requiring a noncompete on page 7 of the contract is to reduce compensation paid by Firm A and Firm B by 12 percent, while the causal effect of a worker signing one is to reduce compensation by 4.2 percent and up to 7.8 percent for those who skipped the noncompete. They find similar effects regardless of whether the noncompete is enforceable.

While these findings are from a specific context (freelance human resources professionals in the finance industry), they directly contradict the efficient contracting theory, which presupposes that workers would never agree to restrictions that hurt them and instead would need to receive compensation for accepting additional restrictions. Instead, Cowgill, Freiberg, and Starr (2024) show that behavioral elements such as contract reading and subjective beliefs are crucial for understanding the results.

A final set of arguments relate to the intersection of training investments and wages. FTC Chairman Andrew Ferguson (2024) made the argument from the efficient contracting view in his dissent from the FTC's proposed noncompete rule:

> Economic theory predicts that, at least in some circumstances, noncompete agreements promote an employer's investment in its employees by mitigating the risk that a rival will ride freely on those investments by luring the employee away before the investing employer can recoup the return on those investments. Investments in employee training and research and development are neither unfair nor anticompetitive. They are critical to the success of employers, employees, and the general economy.

In their comment to the FTC, Gibson, Dunn, & Crutcher LLP (2023) further this argument by emphasizing the potentially deleterious effects for young workers, "for whom on-the-job training and mentorship are indispensable to future success . . ."

While the evidence is mixed on whether noncompetes are associated with more training, the evidence also suggests that workers do not benefit on net from that training.[4] For example, Starr, Prescott, and Bishara (2021) and Starr (2019d) find that greater noncompete enforceability is associated with more training—particularly for workers with noncompetes or in jobs where noncompetes are common—but lower wages. Balasubramanian et al. (2020) use the Longitudinal

[4] Gopal, Li, and Rawling (2025) find no statistically significant effects of noncompetes on formal training. Also, Starr, Ganco, and Campbell (2018) suggest that one reason we see more training in states where noncompetes are enforceable is because noncompetes prevent firms from hiring experienced workers in their industry, thus causing them to hire outsiders or newcomers and train them.

Employer-Household Dynamics data to study the long-run effects on high-tech workers of simply starting a job in a state that enforces versus a state that does not enforce noncompetes. Leveraging the fact that noncompetes are most common among technical workers, they examine how the within-state wage differential between technical workers and nontechnical workers varies when noncompetes are more versus less enforceable. They argue that if the efficient contracting story has merit, then workers who experience training benefits from noncompete enforceability should at some point have higher compensation. Instead, they find that technical workers starting a job in a state that enforces noncompetes experience relative earnings losses, including lower within-individual earnings growth, that lasts over at least eight years. These results suggest that any training benefits accrue primarily to firms, while workers bear the costs of reduced mobility and slower wage growth.

While much literature looks at compensation for the specific workers that sign noncompetes or how market-level wages adjust in states that enforce or ban such contracts, a recent strand of research considers the implications of how the widespread use of noncompetes can spillover to other workers or across state lines. For example, when many firms in an area use noncompetes, then how does that change industry dynamics? Using nationally representative data from the 2014 Noncompete Survey Project, Starr, Frake, and Agarwal (2019) consider this question using a cross-sectional design and find that, where enforceable noncompetes are used collectively, the whole labor market is slower moving and lower earning, including for workers not bound by noncompetes. Johnson, Lavetti, and Lipsitz (2025) construct a dataset of the level of enforceability of noncompetes for each state from 1991 to 2014 and then document cross-state spillovers in how greater noncompete enforceability can suppress compensation both within and across state-boundaries. Gottfries and Jarosch (2023) build a model of wage posting and on-the-job search, calibrate the model based on state-level data, and argue that because of these spillovers "a ban is a relatively cheap way of restoring wage competition"—even considering that a noncompete ban may generate excessive turnover costs that make it inefficient.

Are Noncompetes Necessary Compared to the Alternatives?

Noncompetes are not the only possible contractual tools for protecting legitimate business interests, and for many purposes, narrower and more tailored tools may suffice. In advocating for a bill to ban most noncompetes in Washington, DC, Councilmember Elissa Silverman (2020, p. 2), pointed to "trade secret laws; non-disclosure agreements which legally bind two parties to confidentiality; and non-solicitation agreements which prohibit a former employee from taking clients, customers, or patients from a former employer."

Others strongly contest this view. In a letter urging Governor Kathy Hochul to veto a ban passed by the New York legislature, Russell Beck (2023), a prominent

noncompete attorney wrote that noncompetes are a necessary "prophylactic" tool. As he put it:

> [E]ven where an employee attempts to abide by these other restrictions, the subtle use of trade secrets or indirect solicitation of customers is frequently inevitable simply by dint of the knowledge the employee has or the contact that the employee will have with customers they used to work with. Unlike a breach of a noncompete, this conduct (whether intentional or inadvertent) frequently goes undiscovered until the harm has occurred and is likely irreversible. As a result, none of these alternatives provides the level of protection, deterrence, and clarity offered by noncompete agreements. The elimination of noncompetes will lead to a significant increase in the likelihood that trade secrets will be unlawfully taken to a competitor. Further, a noncompete ban will result in a substantial increase in trade secret litigation as a substitute for noncompete enforcement litigation . . .

Three recent studies speak directly to the questions raised by these claims. The Cowgill, Freiberg, and Starr (2024) field experiment mentioned earlier provides the most direct test. Again, in that study the authors randomized the use of noncompetes in job offers at one firm ("Firm A") and later measured whether workers transferred Firm A's confidential information to a competitor ("Firm B"). While the noncompete reduced mobility to the competitor by 57 percent, it had no effect on the unauthorized sharing of information covered by a nondisclosure agreement. In other words, the findings suggest that workers who violated a nondisclosure agreement would also violate their noncompete, while those who honored the noncompete would have also honored the nondisclosure agreement. As a result, the noncompete simply reduced mobility without adding protection against disclosure of confidential information.

In a study of noncompete bans across the United States, Greenwood, Kobayashi, and Starr (forthcoming) confirm the earlier findings that noncompete bans increase mobility and reduce noncompete litigation, but they also find that trade secret litigation remains flat in the short run and falls in the long run. That is, greater mobility between firms does not seem to increase trade secrecy concerns.

Finally, after Washington's 2020 law banned noncompetes for workers earning less than $100,000, Hiraiwa, Lipsitz, and Starr (2024) test whether firms value noncompetes enough to raise wages to preserve enforceability—that is, do firms respond to the new law by bumping wages just above the threshold, as one might expect if firms viewed the enforceability of noncompetes were critical. The authors find no evidence that wages are bunched in this way, even in technical industries. A contemporaneous survey of the Washington Bar Association found that most attorneys believed companies had adequate alternative tools to protect their interests without noncompetes. Reflecting this sentiment, Microsoft announced in 2022 that it would eliminate noncompetes for all US employees except senior executives (Pannoni and Coleman 2022).

Taken together, these studies suggest that the replacement of noncompetes with alternative legal tools—nondisclosure and nonsolicitation agreements, as well as trade secret protections—appear sufficient to protect firm interests in most cases. In cases where those existing tools are insufficiently protective, there are still other alternatives. For example, firms can directly insure their trade secrets or partner with companies willing to finance trade secret litigation. Firms can also revamp their internal secret-sharing procedures. For example, rotation programs can create multiple ties with clients so that clients do not become attached to a single individual (Battison, Espinosa, and Liu 2025). Firms may also substitute away from trade secrecy toward patenting (Kang and Lee 2022). In terms of contractual alternatives, tailored training repayment agreements may help recover training expenses if workers leave before the firm has recouped their investment (Long 2005; Prescott, Schwab, and Starr 2024). "Garden leave," in which the firm pays the worker to tend their garden and not come to work, can ensure the worker is compensated while preventing them from joining a competitor (Sullivan 2016). Some California firms have worked around the lack of enforceability of executive noncompetes by tying the post-employment payout period to the prohibition period in the noncompete (Sanga 2018). Kräkel and Sliwka (2009) suggest that option contracts can be preferable to a strict noncompete. Similarly, "carrots" like stock options and forfeiture for competition clauses may also prevent workers from joining a competitor (Brown 2023; Aran 2018). Firms may also use fixed term contracts, as some California firms do (*Twentieth Century Fox Film Corp. v. Netflix*, Court of Appeal of the State of California, Second Appellate District Division Five, 2021).[5]

Naturally, these alternatives come with their own tradeoffs (Starr 2022). For example, training repayment agreements can also be thrust upon unsuspecting workers and have inflated repayment penalties, making it very costly for workers simply to leave their job (Harris 2020). Similarly, garden leave, while ensuring that workers are compensated, may inefficiently sideline otherwise productive workers (Azevedo, Pereira, and Rodrigues 2018). Nondisclosure agreements can be written so broadly that some have interpreted them as effective noncompetes (*TLS Mgmt. and Mktg. Ser. LLC v. Rodriguez-Toledo, 19-1104P (1st Cir. [2020])*; Hrdy and Seamans 2024). Ultimately, while alternatives to noncompetes appear broadly sufficient, how firms choose among them—and with what consequences for innovation, mobility, and welfare—remains an important open question.

Do Noncompetes Boost or Stifle Innovation?

Perhaps no issue in the noncompete debate generates more heat than their effects on innovation—a debate often epitomized by the contrast between Silicon

[5] Some suggest that absent noncompetes, companies might simply agree to collude with each other (for example, Ghosh and Shankar 2017). However, the US Department of Justice (2016) clarified in its "Antitrust Guidance for Human Resource Professionals" that such naked "no poach" agreements were illegal.

Valley's open labor markets and the corridor of high-technology firms near Route 128 in Massachusetts, where restrictive labor covenants were common. Critics argue that by suppressing worker mobility, noncompetes stifle the knowledge diffusion, collaboration, and entrepreneurship that drive innovation. Gilson (1999) and Hyde (2003) famously linked Silicon Valley's dynamism to California's 1872 ban on employee noncompetes, while Lobel (2013) emphasized how innovation thrives on "knowledge spillovers" that occur when employees are free to move, start firms, and recombine ideas (Almeida and Kogut 1999). From this perspective, noncompetes are not a catalyst for innovation, but a brake: they impede the flows of knowledge and talent that sustain dynamic innovation ecosystems.

On the other side, supporters argue that noncompetes give firms the confidence to invest in research and development and to share sensitive knowledge with employees, thereby spurring productivity and invention (Barnett and Sichelman 2020). The US Chamber of Commerce, in its comment opposing the FTC's proposed ban on noncompetes, stressed that noncompetes "give firms the confidence to invest in training and research and development" by ensuring that employees cannot immediately transfer valuable knowledge to competitors (see also commentary by Cowen 2023). The intuition is familiar: just as patents and copyright laws allow firms to capture returns on innovation, noncompetes are thought to strengthen incentives for firm-level investment and thus innovation.

While an earlier literature touches on these questions somewhat indirectly or with significant limitations,[6] the recent empirical evidence points toward the following conclusion: While noncompete enforceability modestly increases firm investment, its overall effect is to reduce innovation. The mechanisms appear to include lower worker mobility, fewer startups, weaker knowledge diffusion, reduced employee effort, and a misallocation of inventive talent.

The most comprehensive evidence comes from two recent studies, Johnson, Lavetti, and Lipsitz (2025) and Reinmuth and Rockall (2025), both of which exploit dozens of state-level changes in noncompete enforceability between the 1990s and mid-2010s. Both find that stronger enforceability reduces patenting, including novel and breakthrough inventions, and provide evidence that reduced employee mobility and weaker knowledge spillovers are key mechanisms. Johnson, Lavetti, and Lipsitz (2025) show that this decline is not driven by substitution into trade secrets or by innovation relocating across state borders. They also find that firms increase

[6]For example, using cross-sectional variation in noncompete enforceability, Samila and Sorenson (2011) examine how noncompete policies moderate the effect of venture capital on innovation, firm starts, and employment. Similarly, Conti (2014) finds some evidence that noncompete enforceability spurs firms to pursue riskier research and development projects, though because he does not report event-study plots, it is not clear if these results are driven by pre-trends. Notably, Xiao (2022) finds the opposite result using different measures of innovation in the medical device industry—finding that exploratory innovation falls in the medical device industry after an increase in enforceability. Belenzon and Schankerman (2013) also provide evidence that noncompete enforceability reduces within-state knowledge spillovers. Carlino (2021) finds that Michigan's 1985 policy change, which led to enforcement of noncompetes, increased quality-adjusted mechanical patents. Yet the increase appears 13 years later, making it unlikely to reflect the 1985 reform and more plausibly driven by other factors affecting innovation in the mid-1990s.

intangible investment without raising physical capital, consistent with noncompetes solving a hold-up problem (and consistent with other research looking at the investment margin like Jeffers 2024). In contrast, Reinmuth and Rockall (2025) document that, like innovation, productivity falls with stricter enforceability.

Other studies reach similar conclusions. He (2025) shows that patents produced after an increase in the enforceability of noncompetes are less valuable, suggesting reduced motivation among inventors. Mueller (2023) finds that inventors forced to change industries after an enforceability increase are 30 percent less productive than before, while those who switch industries voluntarily are more productive—evidence that stricter enforcement reallocates inventive talent inefficiently. Finally, using a general equilibrium endogenous growth model, Baslandze and Vardishvili (2022) seek to quantify these channels. Their framework incorporates the various mechanisms—entry restrictions, incentives for firm innovation, reduced knowledge flows, and weaker competition—and concludes that noncompete enforceability results in growth and welfare losses.

Taken together, this body of work suggests that while noncompetes may provide incentives for firm-level investment, their broader effect is to slow innovation and productivity by weakening the mobility and knowledge flows that fuel dynamic, high-growth economies.

Do Noncompetes Help or Hurt Small Businesses?

Many commentators have expressed concern about how noncompete clauses may affect small businesses. One perspective argues that small businesses that have developed valuable confidential information can be especially susceptible to a key employee leaving with that information, potentially "devastating" the business (for example, Corrigan 2023). The US Chamber of Commerce (2023) notes in their comment to the Federal Trade Commission that "noncompetes encourage the development of more viable market entrants" and highlight the testimony of Sam Westgate, a representative of a trade association, that "if non-compete agreements are not allowed for key employees, the revolving door for those employees could eventually force smaller companies out of business . . ." Under this view, noncompetes can help small businesses better protect their investments, thus improving workers incentives both to form startups and to innovate.

The counter-perspective reflects the idea that noncompetes are restraints both on a firm's ability to hire workers with relevant industry-experience and a direct restraint on firm entry. A comment to the FTC (2024) from a physician illustrates these ideas vividly:

> We own a small family practice in urban Wisconsin. I previously was employed by a large healthcare organization and burned out. When I left to star[t] my own business, I was restricted from working close by, by a non-compete. I spent $24,000 [in] legal fees challenging this successfully. . . . Now as a business

> owner for 5 years, we have the opportunity to hire some physician assistants who have been terminated without cause from my prior employer. I am unable to do so because they also had to sign non-competes. I have seen many disgruntled patients who have delayed care because of this.

Conceptually, either of these perspectives might seem plausible. However, the empirical evidence points to the conclusion that noncompetes hurt small businesses. Many studies find that when states are more likely to enforce noncompetes, new firms are less likely to form and that new firms struggle to hire and grow (for example, Jeffers 2024; Marx 2022; Starr, Balasubramanian, and Sakakibara 2018). For example, Balasubramanian et al. (2021) find that bans on physician noncompetes increase the number of practices, practice employment, and survival rates. Johnson, Lavetti, and Lipsitz (2025) show that enforcing noncompetes not only reduces new firm entry but also reduces the job creation rate. They further find that increases in noncompete enforceability reduce innovation among startups.[7] Kang and Fleming (2020), examining the 1996 Florida statute that until recently many regarded as the most vigorous noncompete enforceability policy in the United States (Bishara 2011), find that the Florida law disproportionately benefited large firms versus small firms by increasing the share of establishment entry and employment coming from large firms.

In a recent survey of 312 small business owners by the Small Business Majority (2023), 44 percent of small business owners report that they have been subject to a noncompete that prevented them from starting or expanding their own businesses, while 35 percent report that they have been prevented from hiring an employee because of a noncompete. While this sample may not be representative of all small business owners, it provides specific evidence that underlies the core mechanisms identified in the empirical literature.

Does Including Executives in Bans on Noncompetes Make Sense?

From an efficient contracting perspective, executives may seem like the prototypical example for reasonable use of noncompetes: sophisticated, with outside options and legal teams to review contracts, likely compensated for what they give up, and entrusted with information that could give a competitor a significant advantage. (Another prototypical example might be the asset management industry, where traders are sophisticated and trading strategies are replicable, easy to do clandestinely, and only profitable when competitors do not know exactly what you are doing.) These ideas are reflected in recent policy decisions. After initially proposing a complete ban on noncompetes, the Canadian province of Ontario exempted executives before passing its law in 2021 (as reported by Demeo and Lafontaine

[7]In a contrasting finding, Reinmuth and Rockall (2025) find little evidence of noncompete enforceability affecting new firm applications or new firms' patenting.

2021). In New York state, the Assembly passed a complete ban including executives, but Governor Hochul (2023) vetoed it, explaining that she wanted to allow New York businesses "to retain highly compensated talent" because the companies have "legitimate interests."

Despite the appeal of the efficient contracting view in this case, there are several reasons why a ban on noncompetes for executives is at least cognizable—and indeed some recent theoretical and empirical work points in this direction. These justifications do not rest on disparities in bargaining power or on harms to the executives themselves: rather, the evidence on this point is mixed (for example, Garmaise 2011; Shi 2023; Kini, Williams, and Yin 2021). Instead, bans on executive noncompetes are primarily justified by their third-party effects, alongside the existence of less restrictive protection tools. Having already reviewed the evidence on protective alternatives above, I focus here on arguments and evidence related to third-party harms.

Banning noncompetes on third-party grounds has some history in US law and medicine. The only profession where noncompetes are universally prohibited in the United States is the practice of law. Rule 5.6 of the American Bar Association (ABA) reads:

> A lawyer shall not participate in offering or making:
> (a) a partnership, shareholders, operating, employment, or other similar type of agreement that restricts the right of a lawyer to practice after termination of the relationship, except an agreement concerning benefits upon retirement;

The ABA's comment on Model Rule 5.6 emphasizes the logic behind the rule is that noncompetes can hurt clients: "An agreement restricting the right of lawyers to practice after leaving a firm not only limits their professional autonomy *but also limits the freedom of clients to choose a lawyer*" (emphasis added).[8] Thus, even though lawyers and law firm executives are sophisticated actors with valuable client information and industry know-how, noncompetes are prohibited because of third-party effects on clients.

A similar argument applies to physician noncompetes. The state-level Court of Appeals of Tennessee at Nashville made the comparison explicitly in *Murfreesboro Medical Clinic, P.A. v. David Udom, Tenn* (Tenn. June 29, 2005):

> In analyzing this issue, we see no practical difference between the practice of law and the practice of medicine. Both professions involve a public interest generally not present in commercial contexts. Both entail a duty on the part

[8] For the full text of these rules and comments, see the website of the American Bar Association at https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_5_6_restrictions_on_rights_to_practice/ and https://www.americanbar.org/groups/professional_responsibility/policy/ethics_2000_commission/e2k_rule56/.

> of practitioners to make their services available to the public. Also, both are marked by a relationship between the professional and the patient or client that goes well beyond merely providing goods or services. . . . In both contexts, restrictive covenants have a destructive impact on those relationships. The rules governing other businesses and trades are not relevant to either the legal or medical profession, as both often require the disclosure of private and confidential information such as, in the context of physician and patient, personal medical or family history . . . The right of a person to choose the physician that he or she believes is best able to provide treatment is so fundamental that we can not allow it to be denied because of an employer's restrictive covenant.

The American Medical Association (2014) echoes this concern in its Code of Medical Ethics: "Covenants-not-to-compete restrict competition, can disrupt continuity of care, and may limit access to care." In 2023, the AMA House of Delegates endorsed banning many physician noncompetes, emphasizing that removing them would improve patient access, expand specialist availability, and reduce inequities (as reported by Robeznieks 2023). Several states have since restricted or banned physician noncompetes for precisely these reasons (Fair Competition Law 2025).

A similar logic extends to executives. While Posner, Triantis, and Triantis (2004) were perhaps the first to formalize this idea, several recent papers emphasize third-party harm. For example, Lipsitz and Tremblay (2024) point out that consumers are not at the table when executives bargain with firms over noncompetes. They show that noncompete enforceability raises product-market concentration and posit that executives may accept noncompetes not to protect investments, but to limit future competition, preserving rents for themselves rather than passing them on to consumers as lower prices. In this view, executives are *potential future competitors* who could foster lower consumer prices and more innovation by starting new firms. The noncompete serves as a form of collusion between current and future competitors.

In a search model of the executive labor market, calibrated with real contract data, Shi (2023) formalizes this perspective. She shows that while noncompetes may maximize bilateral surplus between a firm and its executive, they generate "excessive" terms that extract rents from potential future employers and misallocate executive talent. Like Mueller (2023), she finds that executives are displaced from their most productive uses. Her quantitative conclusion: "[T]he optimal policy [for executives] is close to a ban."

Do Consumers Benefit or Suffer from Noncompetes?

What are the downstream effects of noncompete clauses on consumers? The National Retail Federation amicus brief (Weibust et al. 2024) argues that noncompetes "do not harm consumers," reasoning that banning them would raise wages and thus force firms to pass higher costs on to consumers. In her dissent to the FTC's proposed rule, Commissioner Wilson (2023) noted a study suggesting that an

agreement not to enforce noncompetes led brokers to depart their firms, and then consumers who followed these brokers paid higher fees. Federal Trade Commission (2024) reached the opposite conclusion, finding that "non-competes tend to negatively affect competitive conditions in product and service markets." What does the weight of the evidence suggest?

There are several potential mechanisms linking noncompetes (or their enforceability) to consumer outcomes: prices, product quality, and the supply of goods and services. Some harms are direct, while others are indirect.

One example of direct harm arises when consumers rely on an individual service provider, such as a physician. If a physician with a noncompete wants to leave their job, the geographic and temporal restriction in their noncompete will necessitate them leaving the proscribed geographic area or sitting out of the market (simultaneously, the nonsolicitation provisions in the contract may prohibit them from reaching out to prior clients). In that case, patients may lose access to their chosen provider (as reported in Andrews 2019). Similar disruptions can occur in any context where service relationships matter—say, hairstylists, wealth managers, yoga instructors, or auto mechanics. Consumers must then incur search costs and may face lower-quality replacements. In healthcare, exogenously losing a physician increases both mortality and emergency visits by 4 percent, per a study of Medicare physicians who move farther from patients (Sabety 2023). A comparable risk arises when noncompetes sideline the only service provider in an area, depriving local customers of access altogether. Reflecting this concern, a 2019 Florida statute (§542.336) voided noncompetes in cases where one entity employs or contracts with all physicians of a given specialty in a county.

Indirect harms can operate through labor and product markets. If noncompetes suppress wages, banning them might seem to raise costs for firms—with potential price pass-through to customers. But competitive labor markets also expand employment relative to monopsonistic ones, increasing supply in product markets and putting downward pressure on prices. Goudou (2022) models this mechanism, showing that noncompetes reduce employment, while Johnson, Lavetti, and Lipsitz (2025) find that enforceability dampens job creation.[9]

A further channel is market structure. Noncompetes make new firm entry harder and increase incentives for consolidation. Using matched employer-employee data from LinkedIn, Jeffers (2024) documents that noncompete enforceability reduces entry; using a natural experiment in which Michigan legislators attempting to revise an antitrust law managed also to reverse its prohibition on enforcing noncompete clauses, Younge, Tong, and Fleming (2014) document higher merger activity; and using the 1996 change in Florida law regarding noncompetes, Kang and Fleming (2020) find that stricter enforcement tilts growth toward larger incumbents. Concentration, in turn, raises consumer prices and reduces output. Hausman and Lavetti (2021) show this directly in healthcare: enforcing noncompetes leads to higher

[9] Notably, Potter, Hobijn, and Kurmann (2024) reach the opposing conclusion, suggesting that noncompete clauses increase job creation based on their theoretical model of job search.

prices via concentration. Consumers can also be affected if enforcing noncompetes lowers both the quantity and quality of innovation, as discussed earlier.

The financial services sector has provided a natural experiment for some of these issues. Financial firms can sign the "Broker Protocol," which is an established set of rules first laid out in 2004 that (if followed) allow brokers to move freely to a new firm and recruit their prior clients. On one side, Gurun, Stoffman, and Yonker (2021) find that joining the Protocol was associated with increased fees and misconduct. On the other side, Clifford and Gerken (2021), studying the same policy with a broader sample, reach the opposite conclusion: after entry to the Protocol, client disputes fell 20 percent and brokers invested in costly licenses to offer more products—behavior consistent with becoming more client-oriented. Such divergence highlights the need for further research in this specific case to reconcile these findings. In any case, the Broker Protocol should broadly not be interpreted as evidence on noncompetes alone, as it also relaxed nonsolicitation and nondisclosure clauses.

In sum, the preponderance of evidence suggests that noncompetes harm consumers—by disrupting direct service relationships, weakening innovation, reducing entry, and encouraging consolidation that raises prices. The narrative that higher wages from noncompetes would pass through into higher consumer prices is theoretically possible, but empirically less compelling. If anything, more competitive labor and product markets are likely to benefit consumers through lower prices, higher quality, and greater availability.

Is the Status Quo Enforcement Regime Sufficient to Mitigate Harm from Noncompetes?

Now that the national ban on noncompetes proposed by the Federal Trade Commission in 2024 has been blocked by the courts, the regulation of US noncompetes remains under a patchwork of state laws and case-by-case FTC enforcement (Posner 2020). Some observers view this status quo as sufficiently protective, noting that courts are "hesitant" to enforce noncompetes with lower-wage workers (for example, see lawyers Weibust and Gerson 2023). Yet a growing body of research calls the extent of the status quo protections into question.

As Blake (1960, pp. 682–3) observed more than 60 years ago (emphasis added):

> For every covenant that finds its way to court, there are thousands which exercise an *in terrorem* effect on employees who respect their contractual obligations and on competitors who fear legal complications if they employ a covenantor, or who are anxious to maintain gentlemanly relations with their competitors. *Thus, the mobility of untold numbers of employees is restricted by the intimidation of restrictions whose severity no court would sanction.*

In short, the existence of a noncompete may matter as much as, or more than, how courts treat it.

This problem persists. As discussed earlier, firms continue to use noncompetes at similar rates regardless of their enforceability (Starr, Prescott, and Bishara 2021; Rothstein and Starr 2022). Some firms even keep deploying contracts even after courts strike them down. For example, after a Michigan court in 2019 deemed Prudential Security's 100-mile, $100,000 damages noncompete clause unenforceable for security guards, the firm continued to use it (FTC 2023b, pp. 3–4). A firm called Total Quality Logistics has done the same with clauses courts previously held to be overbroad—specifically, an attorney in the case noted (as reported by Hawes 2022): "The big problem with TQL's noncompete is that it's drafted so broadly, everyone knows it's overbroad and won't be enforced as written. And courts have held that it is overbroad and can't be enforced as written. But Ohio has a doctrine that authorizes courts to reform overbroad noncompetes."

Why do firms persist with unenforceable noncompetes? A charitable interpretation is that they are hedging against future policy shifts or trying to invoke another state's law (Glynn 2008).[10] A less favorable interpretation is that they exploit the chilling effect of unenforceable agreements, banking on some share of workers and competitors to comply (Fisk 2001).

Empirical evidence supports the less favorable view. Starr, Prescott, and Bishara (2020) find that noncompetes reduce mobility and redirect worker search away from competitors, even where unenforceable. Workers report turning down offers at similar rates in enforcing and nonenforcing states. In the field experiment discussed earlier in Cowgill, Freiberg, and Starr (2024), simply signing a noncompete reduces mobility and earnings, regardless of enforceability. Prescott and Starr (2024) find that workers often believe their noncompetes are enforceable and that firms actively reinforce this misperception. In nonenforcing states, firms were twice as likely to remind workers about noncompetes after they received competitor offers. An information experiment in the same study showed that correcting workers' beliefs increased openness to competitor jobs—but even informed workers often refrained from moving due to fear of litigation or reputational harm.

Together, this evidence suggests that the status quo regime is insufficient to discipline employer behavior. Many, if not most, noncompetes are unenforceable, yet they continue to generate real effects on mobility and labor market outcomes.

Recent reforms have tried to address these problems. Some states prohibit "choice-of-law" clauses that import the law of noncompete-friendly jurisdictions, reducing forum shopping. Others have adopted bright-line bans on noncompetes

[10] For an example of such maneuvering, after the Supreme Court of California allowed a cross-state noncompete case to be decided in Minnesota in *Advanced Medtronic vs. Bionic* (S097308 [2002]), firms started using noncompetes in California with out-of-state choice of law provisions as way to try to enforce noncompetes in another jurisdiction (Kang and Lee 2022). California then passed Section 925 in 2017 to diminish this practice and passed a revision to California Business and Professions Code Section 16600.5 in 2024 to make it clear that noncompetes are unlawful in California regardless of where they were signed and when. Policymakers in other states are also aware of this forum-shopping approach: Washington state, for example, banned noncompetes for workers earning under $100,000 and required that workers in Washington abide by Washington laws.

for certain workers, combined with private rights of action. In Washington state, for example, workers can sue their employers if noncompetes are imposed below income thresholds, which has led to class actions on behalf of line cooks, flooring installers, and cabinetry installers (Hiraiwa, Lipsitz, and Starr 2024). However, the effect of such reforms on overall noncompete usage remains to be seen. Prior research generally suggests that boilerplate language is surprisingly immune from market forces (Choi, Gulati, and Scott 2017).

Alternative legal approaches that seek to remove noncompetes or severely limit them are also under consideration. The UK Department for Business and Trade (2023), for instance, has proposed capping noncompete duration at three months. Other states, such as Massachusetts, have included language related to "springing" noncompetes. These policies allow for noncompetes to be removed from contracts but let noncompetes be imposed as a remedy when some misconduct has occurred (for discussion, see Beck 2023). Lastly, some are calling for an expansion of antitrust laws to better address the antitcompetitive effects of noncompetes (Posner 2020).

The existing patchwork of state policies poses practical challenges for multistate employers, remote workers, or workers making cross-state moves. This heterogeneity, while useful for empirical analyses and testing, strengthens the case for a greater degree of uniformity. Federal legislation is one possibility. Another is the Uniform Law Commission, a nonpartisan organization that seeks greater uniformity across state laws. It has produced a proposed Uniform Restrictive Employment Agreement Act (Arnow-Richman 2024), which offers model rules for restrictive employment covenants.

Conclusion

A decade ago, Norman Bishara and I observed that the debate over noncompetes was long on theory and short on evidence (Bishara and Starr 2016). Today, dozens of studies have filled in much of that gap. Moreover, in the last decade or so, the large number of new rules passed about noncompetes at different US states and times, as well as in other countries, offers considerable scope for research. How do firms adapt when noncompetes are removed? Which alternatives do they choose, and with what consequences for innovation, mobility, and consumer welfare? Should policymakers address the alternative restrictions to noncompetes? Should rules about noncompetes treat executives differently? What assumptions that noncompetes bring to the fore—such as inattention or mistaken beliefs about enforceability—should supplement the classic efficient contracting model?

Taken as a whole, the evidence from last decade or so supports the conclusion that while noncompetes may create private efficiencies in some narrow circumstances, their widespread use generates net harms that are difficult to remedy under the current patchwork of state laws. Alternatives such as nondisclosure and nonsolicitation agreements appear mostly sufficient to protect firm interests without the broad collateral costs of noncompetes. In short, the emerging consensus is not that

noncompetes never create private gains, but that—on balance—they function as restraints of trade: depressing mobility and wages, misallocating talent, slowing entrepreneurship and innovation, harming third parties, and raising concentration with downstream risks for consumers. The recent empirical evidence on noncompetes has begun to shift the terms of the conversation—and will continue to do so as the next chapter in the noncompete debate unfolds.

References

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

American Medical Association. 2014. "Report of the Council on Ethical and Judicial Affairs." https://code-medical-ethics.ama-assn.org/sites/amacoedb/files/2022-09/11.2.3.1%20Restrictive%20covenants%20–%20background%20reports.pdf.

Andrews, Dan, and Andrea Garnero. 2025. "Five Facts on Non-compete and Related Clauses in OECD Countries." OECD Economics Department Working Paper 1833.

Andrews, Michelle. 2019. "Did Your Doctor Disappear without a Word? A Noncompete Clause Could Be the Reason." *New York Times*, March 15. https://www.nytimes.com/2019/03/15/business/physician-non-compete-clause.html.

Aran, Yifat. 2018. "Beyond Covenants Not to Compete: Equilibrium in High-Tech Startup Labor Markets." *Stanford Law Review* 70: 1235–94.

Arnow-Richman, Rachel. 2019. "The New Enforcement Regime: Revisiting the Law of Employee Competition (and the Scholarship of Professor Charles Sullivan) with 2020 Vision." *Seton Hall Law Review* 50: 1223–59.

Arnow-Richman, Rachel. 2024. "The Uniform Restrictive Agreement Act in a Vast Sea of Strange Reform: An Introduction." *University of Florida Journal of Law and Public Policy* 34: vii–xiii.

Azevedo, Alcino, Paulo J. Pereira, and Artur Rodrigues. 2018. "Non-compete Covenants, Litigation and Garden Leaves." *Journal of Business Research* 88: 197–211.

Balasubramanian, Natarajan, Jin Woo Chang, Mariko Sakakibara, Jagadeesh Sivadasan, and Evan Starr. 2022. "Locked In? The Enforceability of Covenants Not to Compete and the Careers of High-Tech Employees." *Journal of Human Resources* 57 (S): S349–96.

Balasubramanian, Natarajan, Mariko Sakakibara, Evan Starr, and Seethalakshmi Ramanathan. 2021. "Association between Physician Noncompete Agreements and Healthcare Access." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4630026.

Balasubramanian, Natarajan, Evan Starr, and Shotaro Yamaguchi. 2024. "Employment Restrictions on Resource Transferability and Value Appropriation from Employees." *Strategic Management Journal* 45 (12): 2519–47.

Barnett, Jonathan M., and Ted Sichelman. 2020. "The Case for Noncompetes." *University of Chicago Law Review* 87: 953–1049.

Bartelsman, Eric, Sabien Dobbelaere, and Alessandro Zona Mattioli. 2024. "Non-compete Agreements, Tacit Knowledge and Market Imperfections." IZA Discussion Paper 17260.

Baslandze, Salomé, and Ia Vardishvili. 2022. "Born Different: Entrepreneurship through Inventor Mobility, Innovation, and Growth." Unpublished. https://www.dropbox.com/scl/fi/hkmis4gygbkwegfy3k1oi/BornDifferent_Baslandze_Vardishvili.pdf?rlkey=40t8tk8w9uauvvq3j7xcaiclm&e=1&dl=0.

Battiston, Diego, Miguel Espinosa, and Shuo Liu. 2025. "Talent Poaching and Job Rotation." *Management Science* 71 (4): 2975–92.

Belenzon, Sharon, and Mark Schankerman. 2013. "Spreading the Word: Geography, Policy, and Knowledge Spillovers." *Review of Economics and Statistics* 95 (3): 884–903.

Beck, Russell. 2023. "Letter to Governor Kathleen Hochul." Beck Reed Riden LLP, July 5. https://faircompetitionlaw.com/wp-content/uploads/2023/07/NY-20230705-Letter-from-Russell-Beck-to-Governor-Katheen-Hochul-re-Proposed-Noncompete-Ban-CONFORMED-FINAL.pdf.

Bishara, Norman D. 2011. "Fifty Ways to Leave Your Employer: Relative Enforcement of Covenants Not to Compete, Trends, and Implications for Employee Mobility Policy." *University of Pennsylvania Journal of Business Law* 13: 751–95.

Bishara, Norman, and Lorenzo G. Luisetto. 2025. "Goodbye, Noncompetes? Why Now and Policy Convergence." Preprint, SSRN. B http://dx.doi.org/10.2139/ssrn.5345981.

Bishara, Norman D., Kenneth J. Martin, and Randall S. Thomas. 2015. "An Empirical Analysis of Noncompetition Clauses and Other Restrictive Postemployment Covenants." *Vanderbilt Law Review* 68 (1): 1–51.

Bishara, Norman D., and Evan Starr. 2016. "The Incomplete Noncompete Picture." *Lewis and Clark Law Review* 20: 497–546.

Blake, Harlan M. 1960. "Employee Agreements Not to Compete." *Harvard Law Review* 73 (4): 625–91.

Boeri, Tito, Andrea Garnero, and Lorenzo G. Luisetto. 2025. "Noncompete Agreements in a Rigid Labor Market: The Case of Italy." *Journal of Law, Economics, and Organization* 41 (3): 930–58.

Boesch, Tyler, Jacob Lockwood, Ryan Nunn, and Mike Zabek. 2023. "New Data on Non-compete Contracts and What They Mean for Workers." Federal Reserve Bank of Minneapolis, June 21. https://www.minneapolisfed.org/article/2023/new-data-on-non-compete-contracts-and-what-they-mean-for-workers.

Brown, Andrew A. 2023. "Ainslie v. Cantor Fitzgerald, L.P." *Delaware Journal of Corporate Law* 47 (3): 781–83.

Burdett, Kenneth, and Dale T. Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review* 39 (2): 257–73.

Callahan, Maureen B. 1985. "Post-employment Restraint Agreements: A Reassessment." *University of Chicago Law Review* 52 (3): 703–28.

Carlino, Gerald A. 2021. "Do Non-compete Covenants Influence State Startup Activity? Evidence from the Michigan Experiment." Federal Reserve Bank of Philadelphia Working Paper 21-26.

Choi, Stephen J., Mitu Gulati, and Robert E. Scott. 2017. "The Black Hole Problem in Commercial Boilerplate." *Duke Law Journal* 67: 1–77.

Clifford, Christopher P., and William C. Gerken. 2021. "Property Rights to Client Relationships and Financial Advisor Incentives." *Journal of Finance* 76 (5): 2409–45.

CMA Microeconomics Unit. 2024. *Competition and market power in UK labour markets*. Competition & Markets Authority, Report no. 1, January 25. https://assets.publishing.service.gov.uk/media/67531c7221057d0ed56a0422/1._Competition_and_market_power_in_UK_labour_markets.pdf.

Colvin, Alexander J. S., and Heidi Shierholz. 2019. "Noncompete Agreements: Ubiquitous, Harmful to Wages and to Competition, and Part of a Growing Trend of Employers Requiring Workers to Sign Away Their Rights." Economic Policy Institute, December 10. https://www.epi.org/publication/noncompete-agreements/.

Conti, Raffaele. 2014. "Do Non-competition Agreements Lead Firms to Pursue Risky R&D Projects?" *Strategic Management Journal* 35 (8): 1230–48.

Corrigan, John. 2023. "Non-compete Clause Ban Will Have 'Disastrous Effect' on Small Business." Human Resources Director, March 6. https://www.hcamag.com/us/specialization/employment-law/non-compete-clause-ban-will-have-disastrous-effect-on-small-business/438418.

Cowen, Tyler. 2023. "Noncompete Contracts Can Help Both Workers and Firms." *Bloomberg Opinion*, January 10. https://www.bloomberg.com/opinion/articles/2023-01-10/proposed-ftc-ban-on-noncompete-contracts-would-hurt-workers-too.

Cowgill, Bo, Brandon Freiberg, and Evan Starr. 2024. "Clause and Effect: Theory and Field Experimental Evidence on Noncompete Clauses." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.5012370.

Demeo, Matthew, and David Lafontaine. 2021. "Bill 27 Passes, but Not before an Important Amendment to the Ban on Non-competes." DLA Piper, December 3. https://www.dlapiper.com/en/insights/publications/2021/11/bill-27-update-non-compete-agreements-are-not-out-entirely.

Fair Competition Law. 2025. "The Changing Landscape of Trade Secrets Laws and Noncompete Laws Around the Country." September 30. https://faircompetitionlaw.com/changing-trade-secrets-noncompete-laws.

FTC (Federal Trade Commission). 2023a. "Non-Compete Clause Rule (NPRM)." Docket ID FTC-2023–0007. https://www.regulations.gov/docket/FTC-2023-0007.

FTC (Federal Trade Commission). 2023b. "In the Matter of Prudential Security, Inc." Docket No. C-4787, February 23. https://www.ftc.gov/system/files/ftc_gov/pdf/c47872210026prudentialsecurityfinalconsent.pdf.

FTC (Federal Trade Commission). 2024. "Non-Compete Clause Rule." 16 CFR Parts 910 and 912.

Ferguson, Andrew N. 2024. "Dissenting Statement of Commissioner Andrew N. Ferguson Joined by Commissioner Melissa Holyoak: In the Matter of the Non-compete Clause Rule Matter Number P201200." June 28. https://www.ftc.gov/system/files/ftc_gov/pdf/ferguson-noncompete-dissent.pdf.

Fisk, Catherine L. 2001. "Reflections on the New Psychological Contract and the Ownership of Human Capital." *Connecticut Law Review* 34: 765–85.

Garmaise, Mark J. 2011. "Ties that Truly Bind: Noncompetition Agreements, Executive Compensation, and Firm Investment." *Journal of Law, Economics, and Organization* 27 (2): 376–425.

Gibson, Dunn, & Crutcher LLP. 2023. "Comment from Gibson, Dunn & Crutcher LLP." Regulations.gov, May 4, https://www.regulations.gov/document/FTC-2023-0007-21072.

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

Ghosh, Suman, and Kameshwari Shankar. 2017. "Optimal Enforcement of Noncompete Covenants." *Economic Inquiry* 55 (1): 305–18.

Glynn, Timothy P. 2008. "Interjurisdictional Competition in Enforcing Noncompetition Agreements: Regulatory Risk Management and the Race to the Bottom." *Washington and Lee Law Review* 65 (4): 1381–1444.

Gopal, Bhargav, Xiangru Li, and Luke Rawling. 2025. "Do Non-compete Agreements Help or Hurt Workers? Evidence from the NLSY97." Canadian Labour Economics Forum Working Paper 89.

Gottfries, Axel, and Gregor Jarosch. 2023. "Dynamic Monopsony with Granular Firms." NBER Working Paper 31965.

Goudou, Felicien. 2022. "The Employment Effects of Non-compete Contracts: Job Creation versus Job Retention." Preprint, SSRN. https://ssrn.com/abstract=4231219.

Government Accountability Office. 2023. *Noncompete Agreements: Use Is Widespread to Protect Business' Stated Interests, Restricts Job Mobility, and May Affect Wages.* GAO-23-103785.

Greenhouse, Steven. 2014. "Noncompete Clauses Increasingly Pop Up in Array of Jobs." *New York Times*, June 8. https://www.nytimes.com/2014/06/09/business/noncompete-clauses-increasingly-pop-up-in-array-of-jobs.html.

Greenwood, Brad N., Bruce H. Kobayashi, and Evan Starr. Forthcoming. "Can You Keep a Secret? Banning Noncompetes Does Not Increase Trade Secret Litigation." *Journal of Law, Economics, and Organization*.

Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94 (4): 691–719.

Gurun, Umit G., Noah Stoffman, and Scott E. Yonker. 2021. "Unlocking Clients: The Importance of Relationships in the Financial Advisory Industry." *Journal of Financial Economics* 141 (3): 1218–43.

Hardaway, Ayesha Bell. 2016. "The Paradox of the Right to Contract: Noncompete Agreements as Thirteenth Amendment Violations." *Seattle University Law Review* 39: 957–81.

Harris, Jonathan F. 2021. "Unconscionability in Contracting for Worker Training." *Alabama Law Review* 72 (4): 723–83.

Hausman, Naomi, and Kurt Lavetti. 2021. "Physician Practice Organization and Negotiated Prices: Evidence from State Law Changes." *American Economic Journal: Applied Economics* 13 (2): 258–96.

Hawes, Clarissa. 2022. "TQL's Noncompete Hurts Ex-Employee's Job Prospects, Lawsuit Claims." *FreightWaves*, August 31. https://www.freightwaves.com/news/tqls-noncompete-hurts-ex-employees-job-prospects-lawsuit-claims.

He, Zhaozhao. 2025. "Non-compete Agreements, Innovation Value and Efficiency." *Journal of Corporate Finance* 90: 102698.

Hiraiwa, Takuya, Michael Lipsitz, and Evan Starr. 2024. "Do Firms Value Court Enforceability of Noncompete Agreements? A Revealed Preference Approach." *Review of Economics and Statistics*: 1–47. https://doi.org/10.1162/rest_a_01505.

Hochul, Kathy. 2023. "Veto Message—No. 133." https://www.crowell.com/a/web/czSsRVAdepa5GuFvEoV1vu/veto_133.pdf.

Hrdy, Camilla A., and Christopher B. Seaman. 2024. "Beyond Trade Secrecy: Confidentiality Agreements That Act Like Noncompetes." *Yale Law Journal* 133: 669–754.

Hyde, Alan. 2003. *Working in Silicon Valley: Economic and Legal Analysis of a High-Velocity Labor Market*. Routledge.

Jamieson, Dave. 2014. "Jimmy John's Makes Low-Wage Workers Sign 'Oppressive' Noncompete Agreements." *Huffington Post*, October 13. https://www.huffpost.com/entry/jimmy-johns-non-compete_n_5978180.

Jeffers, Jessica S. 2024. "The Impact of Restricting Labor Mobility on Corporate Investment and Entrepreneurship." *Review of Financial Studies* 37 (1): 1–44.

Johnson, Matthew S., Kurt Lavetti, and Michael Lipsitz. 2025. "The Labor Market Effects of Legal Restrictions on Worker Mobility." *Journal of Political Economy* 133 (9): 2735–93.

Johnson, Matthew S., and Michael Lipsitz. 2022. "Why Are Low-Wage Workers Signing Noncompete Agreements?" *Journal of Human Resources* 57 (3): 689–724.

Johnson, Matthew S., Michael Lipsitz, and Alison Pei. 2023. "Innovation, Inventor Mobility, and the Enforceability of Noncompete Agreements." NBER Working Paper 31487.

Kang, Hyo, and Lee Fleming. 2020. "Non-competes, Business Dynamism, and Concentration: Evidence from a Florida Case Study." *Journal of Economics and Management Strategy* 29 (3): 663–85.

Kang, Hyo, and Wyatt Lee. 2022. "How Innovating Firms Manage Knowledge Leakage: A Natural Experiment on the Threat of Worker Departure." *Strategic Management Journal* 43 (10): 1961–82.

Kini, Omesh, Ryan Williams, and Sirui Yin. 2021. "CEO Noncompete Agreements, Job Risk, and Compensation." *Review of Financial Studies* 34 (10): 4701–44.

Kodama, Naomi, Ryo Kambayashi, and Atsuko Izumi. 2025. "Non-compete Agreements and Human Capital Investments." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.5400982.

Kräkel, Matthias, and Dirk Sliwka. 2009. "Should You Allow Your Employee to Become Your Competitor? On Noncompete Agreements in employment Contracts." *International Economic Review* 50 (1): 117–41.

Krueger, Alan B., and Eric A. Posner. 2018. "A Proposal for Protecting Low-Income Workers from Monopsony and Collusion." Hamilton Project Policy Proposal 2018-05.

Lavetti, Kurt, Carol Simon, and William D. White. 2020. "The Impacts of Restricting Mobility of Skilled Service Workers: Evidence from Physicians." *Journal of Human Resources* 55 (3): 1025–67.

Lipsitz, Michael, and Evan Starr. 2022. "Low-Wage Workers and the Enforceability of Noncompete Agreements." *Management Science* 68 (1): 143–70.

Lipsitz, Michael, and Mark J. Tremblay. 2024. "Noncompete Agreements and the Welfare of Consumers." *American Economic Journal: Microeconomics* 16 (4): 112–53.

Lobel, Orly. 2013. *Talent Wants to Be Free: Why We Should Learn to Love Leaks, Raids, and Free Riding*. Yale University Press.

Long, Brandon S. 2005. "Protecting Employer Investment in Training: Noncompetes vs. Repayment Agreements." *Duke Law Journal* 54 (5): 1295–320.

Lucas, Suzanne. 2019. "Warning: Your Internship May Come with a Non-compete Agreement. Remember When Fast Food Was Doing This? This Is a Rotten Business Strategy." *Inc.*, July 5. https://www.inc.com/suzanne-lucas/warning-your-internship-may-come-with-a-non-compete-agreement.html.

Majeed, Omer, Sarah Davies, Nabeeh Zakariyya, Amee McMillan, and Yun Jiang. 2025. "Non-competes: a case of missing wages in Australia." Government of Australia, Department of the Treasury.

Manning, Alan. 2013. *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press.

Marx, Matt. 2011. "The Firm Strikes Back: Non-compete Agreements and the Mobility of Technical Professionals." *American Sociological Review* 76 (5): 695–712.

Marx, Matt. 2022. "Employee Non-compete Agreements, Gender, and Entrepreneurship." *Organization Science* 33 (5): 1756–72.

Meese, Alan J. 2022. "Don't Abolish Employee Noncompete Agreements." *Wake Forest Law Review* 57: 631–709.

Mueller, Clemens. 2023. "Non-compete Agreements and Labor Allocation across Product Markets." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4283878.

Narayan, Ayushi, and Ryan Nunn. 2025. "The Early Labor Market Impacts of Minnesota's 2023 Noncompete Enforcement Ban." Unpublished.

Pannoni, Amy, and Amy Coleman. 2022. "Microsoft Announces Four New Employee Workforce Initiatives." Microsoft blog, June 8. https://blogs.microsoft.com/on-the-issues/2022/06/08/microsoft-announces-four-new-employee-workforce-initiatives/.

Posner, Eric A. 2020. "The Antitrust Challenge to Covenants Not to Compete in Employment Contracts." *Antitrust Law Journal* 83 (1): 165–200.

Posner, Eric A., Alexander Triantis, and George G. Triantis. 2004. "Investing in Human Capital: The Efficiency of Covenants Not to Compete." University of Chicago Law and Economics, Olin Working Paper 137.

Potter, Tristan, Bart Hobijn, and André Kurmann. 2024. "On the Inefficiency of Non-competes in Low-Wage Labour Markets." *Economica* 91 (362): 446–96.

Prescott, J. J., Stewart Schwab, and Evan Starr. 2024. "First Evidence on the Use of Training Repayment Agreements in the US Labor Force." *ProMarket*, March 27. https://www.promarket.org/2024/03/27/first-evidence-on-the-use-of-training-repayment-agreements-in-the-us-labor-force/.

Prescott, J. J., and Evan Starr. 2024. "Subjective Beliefs about Contract Enforceability." *Journal of Legal Studies* 53 (2): 435–88.

Reinmuth, Kate, and Emma Rockall. 2025. "Innovation through Inventor Mobility: Evidence from Non-compete Agreements." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4459683.

Retail Industry Leaders Association. 2024. "Docket No. FTC-2023–0007: Non-Compete Clause Rule, Matter No. P201200" https://www.rila.org/getmedia/764bdbe6-175c-4373-9414-05f8b6e0c4f6/RILA-4-18-23-Comments-to-FTC-on-Proposed-Non-Compete-Rule.pdf?ext=.pdf.

Robeznieks, Andis. 2023. "AMA Backs Effort to Ban Many Physician Noncompete Provisions." American Medical Association, June 13. https://www.ama-assn.org/medical-residents/transition-resident-attending/ama-backs-effort-ban-many-physician-noncompete.

Rothstein, Donna, and Evan Starr. 2022. "Noncompete Agreements, Bargaining, and Wages." *Monthly Labor Review*.

Rubin, Paul H., and Peter Shedd. 1981. "Human Capital and Covenants Not to Compete." *Journal of Legal Studies* 10 (1): 93–110.

Sabety, Adrienne. 2023. "The Value of Relationships in Healthcare." *Journal of Public Economics* 225: 104927.

Samila, Sampsa, and Olav Sorenson. 2011. "Noncompete Covenants: Incentives to Innovate or Impediments to Growth." *Management Science* 57 (3): 425–38.

Sanga, Sarath. 2018. "Incomplete Contracts: An Empirical Approach." *Journal of Law, Economics, and Organization* 34 (4): 650–79.

Saxenian, AnnaLee. 1996. *Regional Advantage: Culture and Competition in Silicon Valley and Route 128, with a New Preface by the Author*. Harvard University Press.

Schwab, Stewart J., and Randall S. Thomas. 2006. "An Empirical Analysis of CEO Employment Contracts: What Do Top Executives Bargain For?" *Washington and Lee Law Review* 63 (1): 231–70.

Shi, Liyan. 2023. "Optimal Regulation of Noncompete Contracts." *Econometrica* 91 (2): 425–63.

Shubber, Kadhim. 2018. "Cushman v the Cleaner: The Fight over Non-competes." *Financial Times*, October 16. https://www.ft.com/content/bfb69d30-ce44-11e8-b276-b9069bde0956.

Silverman, Elissa. 2020. "Report on B23–0494, the 'Ban on Non-compete Agreements Amendment Act of 2020.'" Council of the District of Columbia Committee on Labor and Workforce Development Committee Report, November 19. https://d3n8a8pro7vhmx.cloudfront.net/silverman/pages/1504/attachments/original/1605719108/Committee_Report_B23-494_Ban_on_Non-Compete_Agreements_DRAFT.pdf?1605719108.

Small Business Majority. 2023. "Opinion Poll: Small Business Owners Support Banning Non-compete Agreements." April 13. https://smallbusinessmajority.org/sites/default/files/research-reports/2023-non-compete-poll-report.pdf.

Starr, Evan. 2019a. "Noncompete Agreements and American Workers." Speech, US Senate, November 14, 2019. https://www.sbc.senate.gov/public/index.cfm/2019/11/noncompete-agreements-and-american-workers.

Starr, Evan. 2019b. "Antitrust and Economic Opportunity: Competition in Labor Markets." Speech, US House of Representatives, October 29, 2019. https://docs.house.gov/meetings/JU/JU05/20191029/110152/HHRG-116-JU05-Wstate-StarrE-20191029.pdf.

Starr, Evan. 2019c. "The Use, Abuse, and Enforceability of Non-compete and No-Poach Agreements: A Brief Review of the Theory, Evidence, and Recent Reform Efforts." Unpublished.

Starr, Evan. 2019d. "Consider This: Training, Wages, and the Enforceability of Covenants Not to Compete." *ILR Review* 72 (4): 783–817.

Starr, Evan. 2021. "Are Noncompetes Holding Down Wages?" In *Inequality and the Labor Market: The*

Case for Greater Competition, edited by Sharon Block and Benjamin H. Harris, 127–48. Brookings Institution Press.

Starr, Evan. 2022. "The Ties that Bind Workers to Firms: No-Poach Agreements, Noncompetes, and Other Ways Firms Create and Exercise Labor Market Power." *ProMarket*, January 3. https://www.promarket.org/2022/01/03/workers-poaching-noncompete-employers-labor-antitrust/.

Starr, Evan. 2023a. "Noncompete Clauses: A Policymaker's Guide through the Key Questions and Evidence." Unpublished.

Starr, Evan. 2023b. "Comment by Evan Starr to the Federal Trade Commission Re: Non-compete Clause Rulemaking, Matter No. P201200." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4427741.

Starr, Evan. 2024. "Brief of Small Business Majority and Professor Evan Starr as *Amicus Curiae* in Support of Defendant-Appellant." https://smallbusinessmajority.org/sites/default/files/policy-docs/AS-FILED-SBM%20%2BStarr-Non-Compete-Amicus-CA11.pdf.

Starr, Evan. 2025. "Brief of Professor Evan Starr as *Amicus Curiae* in Support of Appellant." United States Court of Appeals for the Fifth Circuit. https://www.benedictlawgroup.com/_files/ugd/425c89_35f958c311134a64b22651261741ea13.pdf.

Starr, Evan, Natarajan Balasubramanian, and Mariko Sakakibara. 2018. "Screening Spinouts? How Noncompete Enforceability Affects the Creation, Growth, and Survival of New Firms." *Management Science* 64 (2): 552–72.

Starr, Evan, Justin Frake, and Rajshree Agarwal. 2019. "Mobility Constraint Externalities." *Organization Science* 30 (5): 961–80.

Starr, Evan, Martin Ganco, and Benjamin A. Campbell. 2018. "Strategic Human Capital Management in the Context of Cross-Industry and Within-Industry Mobility Frictions." *Strategic Management Journal* 39 (8): 2226–54.

Starr, Evan, J. J. Prescott, and Norman Bishara. 2020. "The Behavioral Effects of (Unenforceable) Contracts." *Journal of Law, Economics, and Organization* 36 (3): 633–87.

Starr, Evan P., James J. Prescott, and Norman D. Bishara. 2021. "Noncompete Agreements in the US Labor Force." *Journal of Law and Economics* 64 (1): 53–84.

Sullivan, Charles A. 2016. "Tending the Garden: Restricting Competition via 'Garden Leave.'" *Berkeley Journal of Employment and Labor Law* 37 (2): 293–325.

Torgnes, Cassandra Lee, and Siri Terjesen. 2025. "Beyond One Size Fits All: A Judicious Perspective on Noncompete Agreements." *Academy of Management Perspectives* 39 (3): 396–420.

UK Department for Business and Trade. 2023. *Smarter Regulation to Grow the Economy*. UK Government.

US Chamber of Commerce. 2023. "Notice of Proposed Rulemaking, Federal Trade Commission; Non-compete Clause Rule (88 Fed. Reg. 3,482–3,546)." https://www.uschamber.com/assets/documents/FTC-Noncompete-Comment-Letter_FINAL_04.17.23.pdf.

US Department of Justice. 2016. "Antitrust Guidance for Human Resource Professionals." October. https://www.justice.gov/atr/file/903511/dl.

US Department of the Treasury. 2016. *Non-compete Contracts: Economic Effects and Policy Implications*. US Department of the Treasury.

Weibust, Erik, and Stuart Gerson. 2023. "FTC's Noncompete Proposal Is Based on Misrepresentations." Portfolio Media, Inc., January 26. https://www.ebglaw.com/wp-content/uploads/2023/01/Gerson-Weibust-January2022-FTCs-Noncompete-Proposal-Is-Based-on-Misrepresentations-Law360.pdf.

Weibust, Erik W., Katherine G. Rigby, Carolyn O. Boucek, and A. Millie Warner. 2024. "National Retail Federation Amicus Brief." https://acrobat.adobe.com/id/urn:aaid:sc:VA6C2:115788a7-b622-4021-9538-3a6e1df78c5c.

White House. 2016. *Non-compete Agreements: Analysis of the Usage, Potential Issues, and State Responses*. White House.

Wilson, Christine. 2023. "Dissenting Statement of Commissioner Christine S. Wilson Regarding the Notice of Proposed Rulemaking for the Non-compete Clause Rule Commission." Federal Trade Commission File No. P201200–1, January 5. https://www.ftc.gov/system/files/ftc_gov/pdf/p201000noncompetewilsondissent.pdf.

Woodman, Spencer. 2015. "Exclusive: Amazon Makes Even Temporary Warehouse Workers Sign 18-Month Non-competes." *The Verge*, March 26. https://www.theverge.com/2015/3/26/8280309/amazon-warehouse-jobs-exclusive-noncompete-contracts.

Xiao, Fenglong. 2022. "Non-competes and Innovation: Evidence from Medical Devices." *Research Policy* 51 (6): 104527.

Young, Samuel. 2024. "Noncompete Clauses, Job Mobility, and Job Quality: Evidence from a Low-Earning Noncompete Ban in Austria." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.3811459.

Younge, Kenneth A., Tony W. Tong, and Lee Fleming. 2014. "How Anticipated Employee Mobility Affects Acquisition Likelihood: Evidence from a Natural Experiment." *Strategic Management Journal* 36 (5): 686–708.

Journal of Economic Perspectives—Volume 40, Number 1—Winter 2026—Pages 167–190

Occupational Licensing in the United States

Janna E. Johnson

Occupational licensing—the requirement to hold a government-issued credential to perform a specific job legally—features prominently in American day-to-day life. Licensed occupations provide haircuts (barbers and cosmetologists), public education (teachers), medical care (physicians and dentists), and home maintenance (plumbers and electricians). Avoiding occupational licensing is nigh to impossible, as the only certain things in life, per the old saying, involve the services of licensed workers: death (funeral directors) and taxes (accountants).

Today, 30 million workers—over one-fifth of the US workforce—require a license to work in their chosen profession, up from just 5 percent in 1950 (Kleiner and Krueger 2010, 2013). The growth of occupational licensure has drawn increasing attention from policymakers over the last decade or so, prompting politicians on both sides of the aisle to wonder if the policy has in some cases gone too far. For instance, a 2015 report published during the Obama administration stated, "Too often, policymakers do not carefully weigh costs and benefits when making decisions about whether or how to regulate a profession through licensing" (US Department of the Treasury Office of Economic Policy, Council of Economic Advisers, and Department of Labor 2015). Along similar lines, the first Trump administration Labor Secretary Alexander Acosta and former South Dakota Governor Dennis Daugaard wrote, "[O]verly burdensome licensure requirements weaken competition without benefitting the public" (Acosta and Daugaard 2018).

■ *Janna E. Johnson is Associate Professor, Humphrey School of Public Affairs, University of Minnesota, Minneapolis, Minnesota.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251458.

Table 1
Fraction of Labor Force with an Occupational License or Certification, and 20 Largest Occupations with Government-Issued Licenses or Certifications

Category	Fraction
Has professional certification or license	0.241
Given has a professional certification or license, fraction reporting it is government-issued	0.900
Largest occupations represented among those holding government-issued certifications or licenses	
Teachers*	0.125
Registered and licensed practical nurses*	0.090
Driver/sales workers and truck drivers	0.030
Lawyers	0.028
Physicians and surgeons*	0.025
Real estate brokers and sales agents*	0.020
Nursing assistants	0.018
Managers, all other	0.018
Barbers, hairdressers, hairstylists, and cosmetologists*	0.017
Construction managers*	0.015
Accountants and auditors*	0.015
Insurance sales agents*	0.013
Education and childcare administrators	0.012
Occupational and physical therapists*	0.012
Securities, commodities, and financial services sales agents, financial managers*	0.012
Electricians	0.011
Medical and health services managers	0.010
Police officers	0.010
Social workers*	0.010
Pharmacists*	0.009
All other occupations	0.501
Observations	273,561

Source: Current Population Survey, January 2023–May 2025.
Note: Sample contains respondents aged 18–64 currently in the labor force and not attending school. Fractions calculated using person-level sampling weights. Occupations marked with an asterisk combine two or more distinct CPS occupation and/or industry codes. For more information, see Supplemental Appendix Table A1.

The ubiquity of occupational licensure is evident in Table 1. Over 24 percent of respondents to the Current Population Survey report holding a professional certification or license, with 90 percent of these issued by the government, the vast majority of these by states (Flood et al. 2024; Gittleman, Klee, and Kleiner 2018). The broad reach of licensure across the economy is illustrated by the 20 largest occupations among those needing a license to work, which comprise both white- and blue-collar occupations (lawyers and truck drivers), high and lower earners (physicians and social workers), and span different sectors of the economy, from health care (nurses and pharmacists), to finance (insurance agents and commodity brokers), to skilled trades (electricians and construction managers). Licensure is highly concentrated among two large occupations—teachers and nurses account for nearly 20 percent of licensed workers—but also affects hundreds of smaller occupations. While the

20 occupations listed in Table 1 account for half of US licensed workers, over 800 occupations are licensed in at least one state (Kleiner 2000).

Advocates for licensure reform often cite the case of florists in Louisiana as an example of unnecessary occupational licensure. The state required licensure of purveyors of cut flower bouquets beginning in 1939—including an exam judging their "skill" in flower arranging (Carpenter 2010). Despite its questionable benefit to consumers in protecting their health and safety—the typical justification for occupational licensure—florists in Louisiana were required to hold a state license until 2024, when it was replaced with a less-restrictive form of occupational regulation following several court challenges (Boehm 2024).

This paper provides background for understanding why reforming occupational licensing policies in the United States is a challenging endeavor, despite bipartisan agreement on its likely necessity. I start with a summary of the economic theory behind occupational licensing and the trade-off faced between providing for public safety and increasing costs to consumers. I then overview the regulatory structure of occupational licensing in the United States and the reasons for its growth, including the strong incentive for existing members of occupations to advocate for tighter licensing rules. I describe what we do and do not know about the impacts of occupational licensure policies on labor markets and quality and safety—highlighting how data challenges have limited research documenting these effects—as well as some other, unanticipated yet potentially important effects of licensure policies. I outline the most commonly proposed licensure reforms, emphasizing that those that have been most widely implemented so far are not necessarily those with the highest potential impact. I conclude by offering some guiding principles and suggestions to those interested in furthering the movement behind both consideration of and research on occupational licensing reform.

The Economics and History of Occupational Licensing

Occupational licensing policies require an individual to earn a license to work for pay in that occupation. To do so, the worker must meet a variety of requirements that vary across occupations, but can include specific education, on-the-job training, and/or passage of licensing exam(s). In this section, I outline the theoretical costs and benefits of occupational licensure, provide a brief primer on the regulatory structure of occupational regulation in the United States, and describe how both consumer-protection motivations and the tendency of occupational licensing to disproportionally benefit certain groups likely contributed to its growth.

The Trade-Off between Cost and Quality

The economic justification commonly given for occupational licensing is that it addresses an asymmetric information problem. If the goods and/or services provided by an occupation require specialized knowledge and training, an average person cannot be expected to evaluate the quality of potential suppliers

effectively, leaving consumers vulnerable to harm caused by low-quality or even negligent provision of these goods and/or services. In theory, requiring licensure protects consumers by ensuring members of the occupation provide goods and/or services that meet or exceed a minimum level of quality. If occupational licensure policies are well-designed, they can increase quality through two potential mechanisms: directly, if the completion of licensure requirements ensures potential members of the occupation acquire skills that increase the quality of the goods and/or services they provide, and indirectly, if the increased costs of entry due to licensing prevent poor-quality workers from joining the occupation (Kleiner and Soltas 2023).

Occupations with the strongest consumer protection arguments in favor of licensure often have both severe asymmetric information problems between providers and consumers and large potential costs to the consumer from poor-quality service. This insight is perhaps most obvious for medical professions, as the average patient would not be able to distinguish between low- and high-quality treatments, and the risk to the patient from poor quality is, in the worst-case scenario, death. Lawyers are another occupation for which quality is hard to discern for the average consumer given the specialized knowledge and training needed to practice law. Although there is no danger of direct physical harm from a bad lawyer as there is from a bad health care provider, hiring a bad lawyer could lead to a false conviction and/or a higher sentence—potentially severe life-altering consequences for the affected client.

Another type of occupation for which licensure may play a key role in protecting consumers is those occupations with the potential for poor quality to only not harm the paying customer, but also the general public. A major motivation for plumber licensure is the prevention of backflow issues from poorly installed fittings, which for systems connected to public water systems could result in outbreaks of water-born illnesses—a potentially hugely costly negative externality. Poor-quality construction or improper wiring in an apartment building could cause the building to collapse or catch fire, leading to the licensing of most construction trades and electricians. The licensure of commercial truck drivers is intended to ensure operators of these vehicles have a minimum level of driving skill, likely preventing the injury and death of others on the road.

However, numerous occupations are licensed in several or many states despite much weaker consumer protection and quality assurance arguments in favor of their licensure. Two such occupations include interior designers (currently licensed in three states) and auctioneers (licensed in 22 states). The presence of a significant asymmetric information problem for these occupations is doubtful, as the average person can likely discern the quality of potential service providers in these occupations themselves relatively easily. An auctioneer candidate could provide a live or recorded demonstration and a potential client can view interior designers' portfolios. (Moreover, interior design quality is quite subjective given the substantial variation in client aesthetic taste.) While an individual could make less money from an estate sale than expected or waste money on bad furniture, the arguments that

licensure of auctioneers and interior designers is critical for consumer protection are less convincing than those for, say, physicians or commercial truck drivers.[1]

Regardless of whether a licensure policy affects quality, it distorts the labor market for that occupation by increasing the cost to enter. If some individuals are prevented from joining the occupation by the licensure cost, the supply of workers in that occupation is reduced. This reduction in labor supply raises prices of the goods and/or services provided by the licensed occupation, increasing the cost to consumers while raising earnings for those in the occupation. This price increase, while costly to consumers, means one key group is very likely to benefit from instituting licensure or increasing its requirements: existing incumbent members of the occupation. Current members of an occupation are usually "grandfathered in" when licensing policies are established or amended, meaning they enjoy the higher wages resulting from licensing's reductions in labor supply without paying the licensure costs. This creates strong incentives for incumbent members of an occupation to push for more stringent, costly licensure requirements than what is justified by quality concerns and to resist any loosening of existing licensure policies.

Whether an occupational licensure policy benefits or harms society overall is largely determined by the size of its effect on quality, demonstrated through the willingness-to-pay framework proposed by Kleiner and Soltas (2023). They characterize the benefits of a licensure policy as hinging on consumers' valuation of its quality effects: if a specific licensure policy raises the quality of goods and/or services provided, consumers may be willing to pay higher prices in return for the quality increase. Indeed, as long as consumers *believe* that a licensure requirement increases quality, they may be willing to pay higher prices, even if the requirement actually does nothing to increase quality! If the increase in consumers' willingness-to-pay offsets the costs of the licensure policy to both consumers (from increased prices) and to new workers (from increased costs to enter the occupation), a specific licensure policy will increase social welfare. However, the Kleiner and Soltas (2023) model does not incorporate any societal benefit from licensure due to preventing negative externalities from poor-quality provision, as these are not necessarily incorporated into consumers' willingness-to-pay. If these benefits are sufficiently large, a licensure policy may still be worth it, even if otherwise its costs would exceed any increase in consumers' willingness-to-pay.

The Complexity of Occupational Regulation in the United States

Which occupations are subject to the "most burdensome" licensure policies? Are some states stricter than others with their approach to occupational licensing? It is tempting to answer these questions by simply counting the number of occupational licensing policies: the more policies in place in a given state or covering a particular occupation, the stricter the licensure. Alas, occupational regulatory

[1] The main argument made for the licensing of all interior designers is that some members of the occupation plan the layout of hospitals and incarceration facilities, two unique cases where interior design may indeed have important health and safety consequences.

structure in the United States is far too geographically complex and variable across occupations for these simple measures to fully characterize differences in occupational licensure across occupations or place.

While over 90 percent of occupational licenses are issued at the state level (Gittleman, Klee, and Kleiner 2018), only considering state licenses will miss county- and city-level licenses common in some occupations. Electricians are licensed in only 37 states, but this does not mean the occupation is completely unregulated in the other 14 states: most, like Pennsylvania and Illinois, instead require electricians to hold county or city licenses, a system also common in other construction trades like plumbers and installers of heating, ventilation, and air-conditioning systems.[2] Some occupations are licensed at the federal level, like airline pilots and ship captains (because they are involved in interstate commerce) and some financial occupations like stockbrokers. It is also not uncommon for an occupation to need licensure at multiple jurisdictional levels simultaneously: some types of financial securities sales agents are both licensed in some states and at the federal level.

Specific state-level requirements for licensure vary considerably across states for some occupations, while for others the requirements are nearly identical everywhere. Lawyers are licensed in every state, but obtaining licensure requires passing the state-specific bar exam, the difficulty of which varies across states. If one simply used state bar exam difficulty to rank the difficulty of entering the legal profession, California would be at the top, but this ranking would miss the fact that California is also one of the few states to allow individuals to obtain a law degree without attending law school (a possibility raised in the public consciousness by reality star Kim Kardashian's pursuit of a law license through this avenue). In contrast, while medical professions like physicians and nurses are licensed at the state level, the major requirement for licensure is identical in all states: passage of one or more national-level licensing exams.

Adding to the complexity of US occupational regulation is the existence of other types of regulation besides licensure. While licensure is both the strictest and most common type—and hence the focus of both this paper and most research on occupational regulation—many occupations are covered by certification and registration policies, sometimes in place of licensure but often in addition to it.

Occupational certification is similar to licensure in that it requires a worker to show a minimum level of competency, such as through passing an exam. However, unlike licensure, lack of certification does not legally bar entry into the occupation. Many occupation certifications are offered by governments, while others are overseen by private-sector organizations. Every state offers a Certified Public Accountant certification, which requires passage of an exam to obtain. While the exact benefits vary by state, and can include the ability to provide some specific accounting services legally (such as attesting to the accuracy of corporate financial

[2] In many areas, construction-related work is covered by business licensing and/or insurance bonding, systems intended to increase consumer protection in addition to or in lieu of the licensing of individual occupation members.

statements), individuals are lawfully allowed to perform other accounting tasks without obtaining the certification. Physicians are licensed at the state level, but most also obtain a national-level certification in a specialty. Physicians who complete the requirements overseen by a specialty board (such as the American Board of Radiology or the American Board of Pediatrics) can refer to themselves as "board certified" in that specialty.

Registration, in contrast to both licensure and certification, does not require any demonstration of competency like completion of education, training, or passing an exam. However, like licensure, registration is required to practice the profession legally (Kleiner and Vorotnikov 2017). To become registered, individuals need only submit their name, address, and qualifications to a regulatory agency, and sometimes complete additional requirements like providing proof of insurance or submitting to a background check. Some occupations are registered in some states but licensed or certified in others, including security guards and pharmacy technicians. Other occupations were historically registered but have since become universally licensed, like real estate agents and insurance brokers. Athletic agents, who recruit and represent college athletes, are registered in many states but otherwise not subject to other forms of regulation.

Both certification and registration are less costly than licensure to workers and consumers, but also have lower potential benefits. While both are legally required, the cost of obtaining a required occupational registration is very low compared with that of completing the education, training, and exam requirements to receive a license, and therefore likely does little to restrict entry into the occupation. In exchange for its low cost, registration offers little protection for consumers from poor-quality providers, aside from perhaps facilitating mechanisms of reputation and accountability through appearing on a public list. The costs to obtain certification are more in line with those for licensure, but unlike registration, workers subject to certification are not required to obtain the credential to enter the occupation. Although certified workers are still subject to competition from uncertified workers, paying the costs of certification may still be worthwhile if it signals a worker's higher quality to potential consumers, enabling them to charge higher prices for their services. However, certification offers a lower level of quality assurance than licensure, as consumers are at risk of harm from poor-quality goods and/or services provided by uncertified workers.

Finally, inconsistences in the language used to describe occupational regulation types further complicates the understanding of differences in credentials across states and occupations. While the definitions I have provided here for licensure, certification, and registration are those used in the economics literature and among those who study occupational regulation, in some cases the terminology used in specific states and/or occupations may not correspond with these definitions. One example is private investigators, who are referred to as "registered" in some states, despite needing to pass an exam and/or meet an experience requirement to provide their services, which would meet the literature's definition of licensure. Moreover, as the differences in these types of regulation are far from common knowledge, the

average consumer likely does not fully grasp the differences in the qualifications required—and the level of quality assurance attached—when a potential provider refers to themselves as "licensed," "certified," or "registered."

The Growth of Licensure in the United States

The expansion of occupational licensing laws in the United States since they became common in the late nineteenth century is attributable to two different causes: states implementing licensing policies for more occupations and shifts in the occupational composition of the labor market from labor market sectors that were typically unlicensed (like manufacturing and farming) towards those in which licensure was more common (like services and health care) (Carollo et al. 2025). By 1950, around 6 percent of the US labor force was licensed, with growth up to that point driven by the expansion of state licensing policies to cover more occupations. Most of these early-licensed occupations were those with the strongest quality-based arguments for licensure, like lawyers and physicians, which continue to be licensed in every state today. However, not all of these early licensure policies are still in effect. For example, commercial photographers were licensed in eight states starting in 1930, but by 1950 all licensure of the occupation was abolished following several court rulings.

Licensure most rapidly expanded during the period from 1950 to 1990, equally driven by states continuing to pass new licensing laws and the shifting composition of the economy. Many of the new licensure policies during this time were not aimed at existing occupations, but rather addressed the creation of new occupations requiring licensure, particularly in the medical field. For example, jobs with a strong quality-assurance case for licensure such as nurse practitioners and physician assistants did not exist prior to 1960, but today are universally licensed in every state. Besides these important exceptions, a good number of new licensure policies passed during this time expanded licensure to occupations with weaker asymmetric information and consumer protection arguments in favor of licensure. However, the additional growth in licensure policies since 1990—from 17 percent of workers to the current 25 percent—is solely attributable to continued shifts in the labor market towards sectors where licensure is more widespread.

How did occupations with weaker social benefit arguments become licensed? The answer likely lies with the relatively high political power held by the group that reaps the highest, immediate benefits from licensure: existing members of the occupation. Many occupations are represented by professional associations, such as the American Medical Association and the American Association of Realtors, as well as state- or local-level organizations, which can either be affiliates of these national bodies or stand-alone organizations. These organizations are funded largely through membership dues paid by existing members of the occupation, with the understanding that one of their roles is to lobby state legislators and other policymakers on behalf of their members. Thus, such organizations have strong incentives to advocate for stricter regulation of their occupations. For example, lobbyists representing interior designers may demonstrate how poorly designed office spaces can

lead to trip hazards, injuring workers. These arguments can have political force, especially because there is usually no organization to present the counterargument against licensure on behalf of consumers and those who might wish to enter the occupation in the future. Carollo et al. (2025) present empirical evidence that the founding of a state-level professional organization leads to a 20 percentage point higher likelihood the represented occupation is licensed in that state within five years relative to states in which no such professional organization exists.

The incentive for incumbent members of an occupation to push for more regulation does not disappear with initial licensure, as tightening of licensure requirements further limits entry into the occupation and raises incumbent wages even more—even if these increased restrictions do nothing to improve quality. Conversely, efforts to lower barriers to entry into the occupation by loosening licensure requirements are likely to face significant pushback from incumbents if they have the potential to lower wages. Consistent with incumbents continuing to push for increasing regulation of their occupations, Carollo et al. (2025) show that in states where an occupation was initially certified or registered, over half of these policies were eventually strengthened to licensure.

Given that an occupation is licensed, how often are its licensure requirements tightened? Have we seen cases where requirements have been loosened besides the handful of cases where licensure of an occupation has been fully abolished? One may think that these questions would be straightforward to answer, although perhaps tedious, by looking up state records of licensure requirements. Sadly, these records do not exist, or certainly not enough of them to trace the history of licensure for most occupations in most states. While initial licensure of occupations is done through legislative statutes, of which states generally keep good historical records, subsequent changes in licensure requirements are overseen and implemented by state licensing boards. These boards have no incentive to keep records of these changes, and the completeness and quality of their records varies both across states and across different licensing boards within a state. In fact, even knowing exact dates of the passage of initial licensing laws is quite difficult, particularly for older occupations, leading to significant variation in these dates across different databases compiled by different groups or researchers (Carollo 2025a).

The Effects of Occupational Licensure

While the discussion and estimation of the effects of occupational licensure policies within economics can be traced all the way back to Adam Smith's *Wealth of Nations*,[3] research on the topic has rapidly expanded over the last few decades. I restrict my focus to this recent work, which uses a variety of modern econometric

[3]Specifically, Smith writes about the restrictions and length-of-service requirements placed on apprentices and on the limitations on the poor moving between parishes (Smith 2016, both passages in Book I, Ch. 10, Part II).

techniques and innovative data sources to reach largely similar conclusions on licensing's effects as earlier work as well as that on licensing in other countries.[4] I also discuss a relatively new area of licensing research that explores indirect or unanticipated effects of the policy.

Effects on Labor Markets

The broad consensus among economists is occupational licensure policies raise the wages of those in the affected occupation relative to what they would be if the occupation were unlicensed, a finding consistent across a variety of contexts and methodologies (examples include Thornton and Timmons 2013; Law and Marks 2017; Gittleman, Klee, and Kleiner 2018; Ingram 2019; Dodini 2023). Kleiner and Soltas (2023) consider occupations with significant between-state variation in the fraction of workers in that occupation that report holding a license (such as security guards, construction managers, and dental assistants) and conclude that fully licensing an occupation leads to 15 percent higher wages than if the occupation was unlicensed. Carollo (2025b) draws a similar conclusion using a different strategy, exploiting the fact that states instituted their initial licensure policies for occupations at different times. He shows that workers in affected occupations in early adopter states have wages 4 percent higher on average 25 years after the introduction of licensure relative to those in later-licensed or unlicensed states.

The empirical findings on the effects of licensure on employment are much less conclusive. Some studies, like Kleiner and Soltas (2023), show licensure leads to substantially lower employment in the affected occupation, while others, like Carollo (2025b), find modest increases in occupational employment following licensure. However, the broad range in estimated employment effects is consistent with economic theory: the effect of an occupational licensure policy on employment depends on the policy's effect on the quality of goods and/or services provided by the licensed occupation. If licensure has no effect on the quality of goods and/or services provided by the occupation, consumers will be unwilling to pay more for them. In this case, the increased cost of entering the occupation from licensing deters individuals from entering and employment in the occupation falls. However, if consumers are willing to pay more post-licensing due to an actual (or perceived) increase in quality, wages rise even more for members of the occupation, encouraging more entrants than if there was no willingness-to-pay effect of licensure. If the amount consumers are willing to pay for the quality increase is large enough, the resulting increase in wages will completely offset the increased cost of entry, and employment in the occupation will increase after licensure. In fact, if the increase in quality generated by a licensure policy is sufficiently large enough relative to the increased cost of entry, its implementation could backfire in the eyes of incumbents,

[4] In addition to Kleiner (2000) in this journal, a review of the historical work on occupational licensing can be found in Kleiner (2006). Kleiner and Koumenta (2022) present a variety of research on licensure's effects across several European countries.

as the number of new entrants drawn into the occupation by the policy could be large enough to reduce the occupation's wages back to the pre-licensure level.

As noted earlier, the size of the willingness-to-pay increase depends on both the size of the quality increase from the specific requirements for licensure as well as the consumers' valuation of the increase in quality for the good and/or service provided by the licensed occupation in question—both of which could vary widely across occupations and licensure standards. As existing studies examine may different occupations and licensing requirements, it is therefore unsurprising that they find a wide variety of employment effects. Furthermore, estimates of labor market effects do not tell us whether licensing an occupation benefits or harms society: to answer this question, we need to know licensure's effects on quality.

Effects on Consumer Safety and Quality

Given the importance of quality effects in the economic argument for occupational licensure, it is perhaps surprising that these effects have received relatively little study compared to licensing's effects on labor market outcomes. However, quality is just much harder to measure than employment or wages! Recent work has tried many creative options, including online reviews of home services providers (Farronato et al. 2024), customer ratings and driving performance of Uber drivers (Hicks et al. 2022), and the quality of financial statements produced by accountants (Posmanick, Chung, and Obie 2024). Others have used more indirect measures of worker quality, such as the selectivity of undergraduate institutions of teachers (Larson et al. 2020), the speed of accountants' career advancement (Barrios 2022), and health care malpractice premiums (Kleiner et al. 2016). As the authors of these studies point out, these metrics are unlikely to measure quality perfectly, and they therefore place significant limitations on their studies' results and their interpretation. Indeed, quality effects for some occupations may not be measurable at all, or at least not within a useful time frame. For example, structural flaws from poor construction may not become apparent unless they lead to failure, which if it occurs may not do so for years or even decades after a building or a bridge is completed.

Regardless of the measure used, nearly all existing work comes to the same conclusion: occupational licensing requirements have little to no effect on (measured) quality. The lack of evidence that licensure achieves its goal of raising the quality of goods and services is cited by those advocating for reduction of or even the complete removal of licensing requirements. They argue that without positive quality effects, licensure creates higher costs for consumers and barriers to entry into the occupation with no benefit to society in return (Edwards 2022; Sweetland and Carpenter 2022). However, like the work on labor market effects, most studies of the quality effects of licensure only examine occupations that are licensed in some states and not others. The occupations that are not licensed in all states likely have weaker quality and consumer protection justifications for licensure in the first place (and perhaps their licensure was indeed driven by incumbent members seeking to raise their wages rather than motivated by protecting the public interest). It is therefore less surprising that studies restricted to these occupations find no effect of licensure on quality.

The lack of evidence that licensure improves quality for occupations with variation in licensure policies across states does not mean that licensure does not raise quality for *all* licensed occupations. The quality effects on licensing for prominent occupations that are universally licensed and have been for over a century—like physicians and lawyers—cannot be measured using between-state comparison techniques, as no such variation exists. One notable exception is Anderson et al. (2020), who show that the adoption of state midwife license laws between 1900 and 1940 reduced maternal death by around 8 percent and may have also reduced infant mortality. The authors were able to use the same state-level policy variation to identify the effects of licensure that studies of more recently licensed occupations employ—almost unheard of for medical occupations—but they had a distinct advantage of having a well-recorded measure of quality (preventing the deaths of mothers and infants). While the positive effects of midwife licensure on the quality of childbirth services Anderson et al. (2020) estimated are not directly applicable to today given the advances in medical care since the early twentieth century, they illustrate a situation in which poor-quality provision significantly raises the chances of severe injury and death—outcomes that customers are likely to be willing to pay relatively high licensure costs to prevent.

Other Effects

A notable expansion in the occupational licensing literature considers indirect or unanticipated impacts of these policies. One strand of this work considers the potential connection between the growth in occupational licensure and some concerning economic trends like declining employment rates, fewer workers changing jobs and employers, and falling rates of interstate migration. The emergence of these trends in the late twentieth century corresponded in time with a rapid increase in occupational licensure (Abraham and Kearney 2020; Austin, Glaeser, and Summers 2018), and as many studies found other demographic and socioeconomic factors explained very little of these shifts (for examples, see Lazear and Speltzer 2012; Molloy, Smith, and Wozniak 2011, 2014, 2017), licensure growth was suggested as a potential cause. However, while evidence so far suggests that licensure reduces labor market fluidity by restricting workers' ability to switch between occupations (Jackson 2023; Kleiner and Xu 2025) and to move between states (Johnson and Kleiner 2020), the magnitudes of these effects—while potentially quite large for affected workers in specific occupations—are too small to explain much of the aggregate trends.

Another emerging area of occupational licensing research considers its role in reducing income gaps across groups. For example, many state licensure policies contain so-called "good moral character" clauses barring convicted felons from holding a license, and thus licensure status can act as a signal of one's nonfelon status. Blair and Chung (2025) show this effect is particularly important for Black men in "ban-the-box" states where employers are barred from directly asking an applicant's felon status: in these states, licensure substantially reduces the difference in wages between Black and non-Hispanic white men. Other studies suggest that licensure

may play a wider role in reducing labor market discrimination for other minority groups, showing licensure reduces differences in wages between minority groups and non-Hispanic whites for nurses (Law and Marks 2017) and dental assistants (Xia 2021). However, Dodini (2023) documents a spillover effect of licensure that may increase earnings inequality: licensing reduces wages in unlicensed occupations with similar skills to licensed occupations, and the effect is largest for women, Black, and foreign-born Hispanic workers.

Reforming Occupational Licensing

I now turn to discussing the three most common types of occupational licensing reform that have been proposed and/or implemented in the past few decades. Keep in mind that the most effective occupational licensing reforms will increase the supply and lower prices of the goods and/or services provided by licensed occupations, have no effect on their quality provision, and impact a large number of licensed workers.

Scope of Practice

One type of licensing reform removes legal restrictions on the specific tasks an occupation can perform, often called "scope of practice." Such restrictions are most common in the medical field. For example, nurse practitioners hold advanced nursing degrees, and their training includes diagnosis and treatment of common medical conditions. However, in many states, nurse practitioners cannot practice without the oversight of a physician, and they and other nonphysician occupations like psychologists are restricted or prohibited from prescribing medications to patients. Several states have recently removed the oversight requirement and/or expanded the prescription authority of these occupations, two examples of licensure reforms removing limitations on occupations' scopes of practice (Norris et al. 2024).

Scope of practice reforms have potentially large social welfare benefits. Granting nurse practitioners the ability to practice without the supervision of a physician does not directly affect the supply of nurse practitioners, as the licensure requirements to become a nurse practitioner remain unchanged. Instead, expanding nurse practitioners' scope of practice increases the supply of basic primary care services like annual checkups and the treatment of common infectious diseases by allowing nurse practitioners to operate their own clinics. This increased availability of basic medical care can improve social welfare by lowering prices of these services, and/or by expanding access to geographic areas and/or populations where availability of such services was limited prior to the reform.

Studies on scope of practice reforms show evidence of benefits to both individual consumers and the public at large. Prices of well-child visits are lower in states that allow nurse practitioners to provide these basic services without physician supervision relative to states that require it (Kleiner et al. 2016). Rates of

induced labor and cesarean section births—two childbirth interventions known to come with health risks to both the child and the mother—are lower in states that allow certified nurse midwives (a type of nurse practitioner specializing in women's health) to provide more maternal health services (Markowitz et al. 2017). Granting prescription power to psychologists (a privilege previously only held by psychiatrists, who are licensed physicians) even appears to improve public health by positively impacting self-reported mental health and reducing suicide (Alexander and Schnell 2019; Choudhury and Plemmons 2023).

However, like any reform, the overall net effect of scope of practice reforms on social welfare hinges on their impact on the quality of services provided. If nurse practitioners in independent practice provide lower-quality primary care services than physicians, the benefits of improved access for consumers is at least partially negated by the lower quality. Consumers could be compensated for the lower quality by paying lower prices for the service, but if the quality difference is large enough, social welfare could be overall unaffected by scope of practice reforms or even decline. Fortunately, the scope of practice reforms implemented to date have not appeared to reduce quality: those studies showing increased access to health care services from expanding nurse practitioner and certified nurse midwife scope of practice do not show any change in health outcomes like infant mortality rates (Kleiner et al. 2016; Markowitz et al. 2017).

Scope of practice reforms may face resistance from occupations that compete with occupations with newly expanded scopes of practice, as their wages may decline following the reform. For example, while expanding nurse practitioner scope of practice allows them to potentially earn higher wages by providing more services, wages of physicians, whose scope of practice was previously "protected," may decline as they face more competition for primary care than before the reform. However, if physicians replace the primary care services they provided prior to the reform with more expensive, specialized services that nurse practitioners cannot provide, physician wages may remain the same after the reform or even rise. Empirical work on how scope of practice reforms affect wages does appear to show a transfer of earnings from the higher-paid occupation to the lower-paid: nurse practitioner and dental hygienist wages increase after their scope of practice is expanded, while those of physicians and dentists decline (Kleiner and Park 2010; Kleiner et al. 2016; Dillender et al. 2023).

Universal License Recognition

The most widely implemented occupational license reform in recent years addresses a cost of occupational licensure created by the US state-level system of occupational regulation: if workers in a licensed occupation want to move between states and remain in their occupation, they must transfer their license to their new state. The cost of license transfer varies dramatically, from simply filling out some paperwork and paying a nominal fee (as is common for most medical occupations), to passing new licensure exams (as most lawyers must), to completing additional years of education or on-the-job training (teachers). Often, the requirements are

highly dependent on the relative initial licensure requirements in one's origin and destination states (like number of required training hours for cosmetologists), and in some cases, the requirements for transfer are determined on an individual-level basis, known as "board discretion." Under board discretion, it is typical for licensing boards to refuse to determine an individual's transfer requirements until they can prove they are resident in their new state.

Public awareness of this issue grew in the early 2010s over concern for military spouses, who can be disproportionately affected by re-licensure costs given their frequent interstate moves when their spouses are transferred between posts (Executive Office of the President 2013). In response, states began to implement policies now collectively known as "universal license recognition," which ostensibly automatically recognize existing all out-of-state occupational licenses as valid in the enacting state. As they were often touted as a way to reduce skilled-labor shortages and increase economic activity, enactment of these policies spread rapidly: since the first implementation in Arizona in 2019, 21 additional states adopted universal license recognition policies by 2024 (Bae and Deyo 2024; Springman 2025).

While the rapid spread of universal license recognition policies is often cited by occupational licensing reform advocates as evidence of its effectiveness in lowering licensure costs for a large number of workers, the potential of these reforms to raise aggregate social welfare in a significant way is limited. First, licensure transfer costs have been shown to limit the interstate migration of only a few licensed occupations with the highest re-licensure costs. Johnson and Kleiner (2020) compare the interstate migration of 14 licensed occupations with high between-state transfer costs to 6 licensed occupations with low costs. Only 5 of the 14 high-cost occupations appear to be limited in their ability to move between states: teachers, lawyers, pharmacists, electricians, and pest control workers. In addition, consistent with re-licensure costs playing little role in the migration decisions of nurses (who, like physicians, face very low re-licensure costs), DePasquale and Stange (2016) find no evidence that the adoption of the Nurse Licensure Compact (intended to ease nurse licensure transfer between states) increased their migration.

Second, despite claiming to reduce or eliminate re-licensure costs for all occupations, most universal license recognition states will only accept out-of-state licenses from states with "substantially equivalent" licensure requirements and/or occupation scope of practice. Such caveats are intended to protect consumers by ensuring licenses are only granted to workers from other states whose skills meet the receiving state's standards for initial licensure in the occupation, but their inclusion in the policy reduces the number of workers for whom universal license recognition actually reduces transfer costs. Moreover, 16 of the 26 states with universal license reform policies as of May 2024 exempt teachers and lawyers from the policy (Bae and Deyo 2024). Explicitly exempting two of the largest occupations with the highest re-licensure costs from universal license recognition policies likely limits their effectiveness significantly, meaning their potential to address skilled-worker shortages and raise economic opportunity may be overstated. Indeed, the inclusion of these caveats and exemptions may have enabled their adoption, as they also likely

reduce the potential for universal license recognition to reduce wages for incumbent members of occupations.

However, despite the low chance that universal license recognition policies improve aggregate social welfare, they are likely still worth implementing. The few studies so far of the effects of the policies link their adoption to increased employment of licensed workers and overall in-migration (Deyo and Plemmons 2022; Bae and Timmons 2023), although more work is needed to prove these increases are caused by the policy. Even if aggregate effects are small or negligible, utility gains from universal license reform for affected individuals could be very large. For example, the ability to transfer licenses easily between states may allow a social worker married to a member of the military to continue their career following their spouse's change in deployment, when before the reform high re-licensure costs may have led that person to leave the labor force instead.

Deregulation

The reduction or removal of occupational license requirements, referred to as deregulation, can cover a range of possible reforms: modest changes in requirements, like reducing the number of training hours required for licensure; replacing an occupation's licensure with a less restrictive form of regulation like certification or registration; or even abolishing all regulation for a given occupation. The social welfare effects of these reforms are highly dependent on their design—remove the "right" requirement and reduce prices of critical goods and/or services, take away the "wrong" one and cause severe harm by allowing unqualified workers to provide these goods and/or services, potentially resulting in injury or even death.

Studies of the effects of historical changes in occupational regulation could help identify the most potentially beneficial types of deregulation reforms, but the information required to perform these studies is a combination of either rare, unique, or unknowable. As complete deregulation of an occupation must occur through statute, we know of a few examples of complete de-licensure of single, small occupations in a handful of states (Thornton and Timmons 2015), but such instances are very rare. Moreover, as most occurred decades ago and affected small occupations, some of which practically no longer exist (such as watchmakers in Minnesota and Wisconsin), determining their effects on prices and quality are unlikely to be very relevant to current discussions of occupational deregulation. There may well be more instances of this type of deregulation, but as policies that have been repealed may not appear in statute records, we will likely never know for sure (Carollo 2025a). Probably more common were either raising or lowering of licensure requirements by licensing boards for existing licensed occupations, but as I discussed previously, records of these changes are at best likely incomplete and in many cases nonexistent.

While we are much more likely to have complete records of recent licensure requirement changes, their full welfare effects may not be measurable until years or even decades later. For example, many states reduced the required training hours for cosmetology licenses during the 2010s from an average of 1,800 to 1,500 hours. A study comparing affected states to those states that did not change their training

requirements shows that reducing training hours increases enrollment and completion among students of cosmetology schools (Rebolledo, Blanchard, and Cellini 2025), indicating a potential increase in the supply of newly-trained cosmetologists. However, it will take some time before we can know the effects of this deregulation on the supply and wages of cosmetologists, as it takes up to a decade or longer for the full impacts of licensure requirement changes affecting new entrants to affect overall labor market outcomes (Han and Kleiner 2021). Even once the labor market effects have materialized, there is still the significant challenge of measuring any quality effects of reduced training hours, given the highly subjective nature of measuring cosmetology service quality (like the quality of a haircut) and the relative rarity of extremely negative cosmetology-related events (like burns from misapplied chemicals or cuts from dropped scissors).

Deregulation reforms likely to have the highest potential benefit are those that remove or reduce licensure requirements that are very costly to fulfill for potential entrants, but are unlikely to have much effect on the quality of goods and/or services provided to consumers and, ideally, impact a large occupation. An example of a deregulation reform that likely meets these criteria is the elimination of the United States Medical Licensure Examination Step 2 Clinical Skills (CS) exam in 2021 (Murphy 2021). Meant to test students on their clinical and communication skills—arguably very important for the effective practice of medicine—the Step 2 CS exam was broadly considered by medical educators to be a poor assessment of these skills. The exam was graded pass/fail with a pass rate over 90 percent and was next to useless as a screening tool for future employers, yet passage was still required for medical students to continue their training. Thus, students had to study to pass the exam, and as it was only offered in person in five cities, it required many students to spend thousands of dollars on travel in addition to the $2,000 examination fee. While elimination of the Step 2 CS exam significantly reduced the cost of licensure for medical students, it is still a small reduction in the total cost of becoming a physician and is therefore unlikely to affect the supply of physicians or the cost of their services. However, as it likely did not affect the quality of future physicians, as the skills it assessed are now done so as part of medical students' practical training, the fact that attaining a medical license no longer requires spending thousands of dollars and hours studying to pass a useless exam is an increase in overall social welfare.

Deregulation proposals are likely to face more pushback from incumbent members of occupations the more likely they are to lower licensure requirements that truly limit the supply of workers in the occupation. The elimination of the Step 2 CS exam was *encouraged* by the American Medical Association, which represents the interests of physicians, probably because the supply of new physicians is more limited by the number of medical school spots available than by medical licensure requirements. A deregulation reform that could increase competition for incumbents from new workers would be if more states instituted an alternative pathway to attain a law license that did not require graduating from law school, such as that available in California. If instituting this pathway lowered the cost of becoming a

licensed lawyer,[5] it could increase their supply and therefore lower wages of existing lawyers. Such a change would certainly be lobbied against by the state's bar association, which would likely argue that allowing individuals to become lawyers without attending law school would open consumers up to potential harm from low-quality legal services. These arguments may indeed have merit, but whether or not the total cost of this harm is greater or less than the value to consumers of lower prices for legal services is very hard to predict.

The Future of Occupational License Reform

Both economic theory and research support the growing concern that perhaps at least some occupational licensure policies have become stricter than needed, increasing the wages of licensed occupations by forcing consumers to pay unnecessarily high prices for their services. However, determining which specific occupations and licensing requirements to target to maximize the effectiveness of potential reforms without needlessly endangering consumers is extremely difficult. While advocates of licensure reform tout the recent rapid spread of universal license recognition as significant progress in lowering licensure's burden, momentum behind occupational licensure reform should not stop there, as other types of reforms likely have higher potential to increase overall social welfare. I conclude by highlighting key concepts to keep in mind when evaluating the usefulness and feasibility of potential reforms, as well as offering some suggestions on how to ensure recent advancements in occupational licensing research continue.

First, the focus of reform efforts should be guided by their potential to increase social welfare. Given the evidence that recent scope of practice reforms for medical professions have increased the supply and/or reduced prices for medical care with no detectible increase in adverse events, and may have improved broader public health outcomes, expanding their implementation to cover more occupations and states could be a productive area of concentration for licensure reform advocates. They could even be expanded outside of medical fields, such as by allowing paralegals to provide basic legal services without the supervision of a lawyer.

Second, the higher the potential for a reform to reduce "burdensome" licensure requirements, the stronger the likely resistance from incumbent occupation members. As such reforms are potential threats to their members' earnings, professional organizations will likely present policymakers with convincing arguments that enacting the reform will harm consumers. These concerns should not be completely dismissed out of hand, as existing members of occupations are likely the best judges of which licensure requirements protect consumers. However, the veracity of these

[5] In practice, very few individuals become licensed lawyers in California through the alternative pathway, as it is considered to be equally or more costly than attending law school given that it requires years-long apprenticeships in addition to passing the notoriously difficult bar exam.

claims should be taken with a grain of salt given the strong incentive for members of occupations to ensure proposed reforms fail. Licensing reform advocates should be prepared to counter these arguments as well as to fight any court challenges, given that one case involving dentists accusing teeth whiteners of practicing dentistry without a license recently made it all the way to the Supreme Court (*North Carolina Bd. of Dental Examiners v. FTC*, 574 US 494 [2015]).[6]

Third, while the design of potential licensure reforms should be guided by empirical evidence when possible, extreme caution is necessary when attempting to generalize research results to other contexts. This sort of exercise is challenging when considering any type of policy change, but is particularly so for licensure polices due to their vast heterogeneity across multiple dimensions. Existing work on occupational licensing often takes the form of case studies of a single or handful of licensed occupations, and the results of one case study are rarely directly translatable to others. Are the wage effects of scope of practice expansions for nurse practitioners likely to be similar to the effects for physician assistants? Is the effect of a reduction in cosmetologist training from 2,000 to 1,500 hours the same as a reduction from 2,500 to 2,000 hours? Other studies combine licensed workers from many different occupations to estimate the effect of a single type of policy reform, which masks potentially important variation in effects of the policy across different occupations. If such a study shows that universal license recognition increases the average migration of all licensed workers, is the effect similar for social workers, insurance agents, and truck drivers, or are only some occupations affected? Moreover, economic research provides very little guidance on occupations with the longest history of occupational licensure, as these occupations are usually excluded from studies of licensure reforms.

Fourth, the expansion of licensure laws to previously unlicensed occupations does not necessarily place an "unnecessary burden" on workers. New occupations with strong consumer protection arguments for licensure continue to emerge, particularly in the medical field. Two such occupations, genetic counselors and diagnostic medical sonographers, are currently licensed in 30 and 4 states, respectively, with other states considering enacting licensure policies. Counterintuitively, licensing more occupations can reduce licensure requirements if these occupations were previously covered by another licensure policy with higher or additional requirements. State licensing of estheticians (specialists in skincare services) and hair braiders (who provide traditional hairstyles to specific types of natural hair) falls into this category, as both occupations were previously covered in many states by cosmetology licensure. Obtaining a cosmetology license required estheticians and hair braiders to be trained in haircutting and dyeing, services which neither occupation provides to customers. Policymakers should not be afraid of increasing the number of licensed occupations in their state if the increase arises through one of these mechanisms.

[6]The ruling was 6–3 in favor of the teeth whiteners and set an important precedent limiting the authority of licensing boards made up of active members of the regulated occupation.

Finally, effort should be dedicated to increasing the usefulness of existing datasets for the study of occupational licensing. There is little doubt that the recent rapid growth in economic research on occupational licensing has been made possible by both the availability of new databases of state-level, occupation-specific licensure information and the addition of licensure questions to federal surveys like the Survey of Income and Program Participation and the Current Population Survey. Unfortunately, the latter appear to suffer from significant measurement error (Carollo 2025a), potentially limiting their usefulness in studies of occupational license reform.[7] A potential solution to the licensure measurement error problem would be the creation of datasets that combine administrative records on occupational credentials with individual-level earnings and employment information. One way to do this is to add occupation and worker licensure status to the Longitudinal Employer-Household Dynamics data, a restricted-use dataset made available to researchers via the secure Census Bureau Research Data Centers. Further research made possible by the creation of datasets like this can help sustain the current momentum behind occupational license reform.

■ *I am grateful to Nicholas Carollo for his invaluable guidance and feedback. Ethan Ellis provided excellent research assistance. I thank Timothy Taylor and his colleagues on the JEP editorial staff for their tremendous help in preparing this essay. I acknowledge the foundational role Morris Kleiner has played in occupational license research in economics, and convey my profound gratitude to him not only for his comments on this paper but also his mentorship and support in my own career. My brother, Dr. Mark Johnson, MD, provided the example of the elimination of the Step 2 CS physician licensure exam. Finally, on behalf of all of us who study occupational licensing, I pay homage to the countless licensed workers from whom we draw inspiration and collect anecdotes, including our own personal care and home service providers, friends and neighbors, and, in my case, my own husband, a public high school teacher.*

References

Abraham, Katherine G., and Melissa S. Kearney. 2020. "Explaining the Decline in the US Employment-to-Population Ratio: A Review of the Evidence." *Journal of Economic Literature* 58 (3): 585–643.

Acosta, Alexander, and Dennis Daugaard. 2018. "Make It Easier to Work without a License." *Wall Street Journal*, January 8. https://www.wsj.com/articles/make-it-easier-to-work-without-a-license-1515457813.

[7]For example, survey respondents who report their occupation is "physician," "lawyer," or "pharmacist" should all also state that they hold a government-issued license, as these occupations are licensed in every state. However, in the sample I use in Table 1, only 88, 83, and 67 percent do, respectively.

Alexander, Diane, and Molly Schnell. 2019. "Just What the Nurse Practitioner Ordered: Independent Prescriptive Authority and Population Mental Health." *Journal of Health Economics* 66: 145–62.

Anderson, D. Mark, Ryan Brown, Kerwin Kofi Charles, and Daniel I. Rees. 2020. "Occupational Licensing and Maternal Health: Evidence from Early Midwifery Laws." *Journal of Political Economy* 128 (11): 4337–83.

Austin, Benjamin, Edward Glaeser, and Lawrence Summers. 2018. "Jobs for the Heartland: Place-Based Policies in 21st-Century America." *Brookings Papers on Economic Activity* 49 (1): 151–255.

Bae, Kihwan, and Darwyyn Deyo. 2024. *Policy Brief: 2024 Update to the Survey of Universal Licensing Reforms in the United States*. West Virginia University Knee Regulatory Research Center.

Bae, Kihwan, and Edward J. Timmons. 2023. "Now You Can Take It with You: Effects of Occupational Credential Recognition on Labor Market Outcomes." West Virginia University Economics Faculty Working Paper 3–2023.

Barrios, John M. 2022. "Occupational Licensing and Accountant Quality: Evidence from the 150-Hour Rule." *Journal of Accounting Research* 60 (1): 3–43.

Blair, Peter Q., and Bobby W. Chung. 2025. "Job Market Signaling through Occupational Licensing." *Review of Economics and Statistics* 107 (2): 338–54.

Boehm, Eric. 2024. "Louisiana Finally Fixes America's Dumbest Licensing Requirement." *Reason*, May 30. https://reason.com/2024/05/30/louisiana-finally-fixes-americas-dumbest-licensing-requirement/.

Carollo, Nicholas A. 2025a. "Historical Data on Occupational Regulation in the United States." Unpublished.

Carollo, Nicholas A. 2025b. "The Labor Market Effects of Occupational Licensing Policies." Unpublished.

Carollo, Nicholas A., Jason F. Hicks, Andrew Karch, and Morris M. Kleiner. 2025. "The Origins and Evolution of Occupational Licensing in the United States." NBER Working Paper 33580.

Carpenter, Dick M., II. 2010. *Blooming Nonsense: Experiment Reveals Louisiana's Florist Licensing Scheme as Pointless and Anti-competitive*. Institute for Justice.

Choudhury, Agnitra Roy, and Alicia Plemmons. 2023. "Effects of Giving Psychologists Prescriptive Authority: Evidence from a Natural Experiment in the United States." *Health Policy* 134: 104846.

DePasquale, Christina, and Kevin Stange. 2016. "Labor Supply Effects of Occupational Regulation: Evidence from the Nurse Licensure Compact." NBER Working Paper 22344.

Deyo, Darwyyn, and Alicia Plemmons. 2022. "Have License, Will Travel: Measuring the Effects of Universal Licensing Recognition on Mobility." *Economics Letters* 219: 110800.

Dillender, Marcus, Anthony T. Lo Sasso, Brian J. Phelan and Michael Richards. 2023. "Occupational Licensing and the Healthcare Labor Market." NBER Working Paper 29665.

Dodini, Samuel. 2023. "The Spillover Effects of Labor Regulations on the Structure of Earnings and Employment: Evidence from Occupational Licensing." *Journal of Public Economics* 225: 104947.

Edwards, Chris. 2022. "Occupational Licensing." In *Empowering the New American Worker: Market-Based Solutions for Today's Workforce*. Cato Institute.

Executive Office of the President. 2013. *The Fast Track to Civilian Employment: Streamlining Credentialing and Licensing for Service Members, Veterans, and Their Spouses*. White House.

Farronato, Chiara, Andrey Fradkin, Bradley J. Larson, and Erik Brynjolfsson. 2024. "Consumer Protection in an Online World: An Analysis of Occupational Licensing." *American Economic Journal: Applied Economics* 16 (3): 549–79.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Backman, Annie Chen, et al. 2024. *IPUMS CPS: Version 12.0*. IPUMS Center for Data Integration. https://doi.org/10.18128/D030.V12.0 (accessed June 16, 2025).

Gittleman, Maury, Mark A. Klee, and Morris M. Kleiner. 2018. "Analyzing the Labor Market Outcomes of Occupational Licensing." *Industrial Relations* 57 (1): 57–100.

Han, Suyoun, and Morris M. Kleiner. 2021. "Analyzing the Influence of Occupational Licensing Duration and Grandfathering on Labor Market Outcomes." *Industrial Relations* 60 (2): 147–87.

Hicks, Jason F., Jonathan Hall, Morris M. Kleiner, and Rob Solomon. 2022. "Occupational Licensing of Uber Drivers." In "Essays on Historical and Contemporary Occupational Licensing in the United States." PhD diss., University of Minnesota.

Ingram, Samuel J. 2019. "Occupational Licensing and the Earnings Premium in the United States: Updated Evidence from the Current Population Survey." *British Journal of Employment Relations* 57 (4): 732–63.

Jackson, Osborne. 2023. "Occupational Licensing and Occupational Mobility." Federal Reserve Bank of Boston Working Paper 23–18.

Johnson, Janna E., and Morris M. Kleiner. 2020. "Is Occupational Licensing a Barrier to Interstate Migration?" *American Economic Journal: Economic Policy* 12 (3): 347–73.

Kleiner, Morris M. 2000. "Occupational Licensing." *Journal of Economic Perspectives* 14 (4): 189–202.

Kleiner, Morris M. 2006. *Licensing Occupations: Ensuring Quality or Restricting Competition?* W.E. Upjohn Institute for Employment Research.

Kleiner, Morris M., and Maria Koumenta, eds. 2022. *Grease or Grit? International Case Studies of Occupational Licensing and Its Effects on Efficiency and Quality*. W.E. Upjohn Institute for Employment Research.

Kleiner, Morris M., and Alan B. Krueger. 2010. "The Prevalence and Effects of Occupational Licensing." *British Journal of Industrial Relations* 48 (4): 676–87.

Kleiner, Morris M., and Alan B. Krueger. 2013. "Analyzing the Extent and Influence of Occupational Licensing on the Labor Market." *Journal of Labor Economics* 31 (S1): S173–202.

Kleiner, Morris M., Allison Marier, Kyoung Won Park, and Coady Wing. 2016. "Relaxing Occupational Licensing Requirements: Analyzing Wages and Prices for a Medical Service." *Journal of Law and Economics* 59 (2): 261–91.

Kleiner, Morris M., and Kyoung Won Park. 2010. "Battles among Licensed Occupations: Analyzing Government Regulations on Labor Market Outcomes for Dentists and Hygienists." NBER Working Paper 16560.

Kleiner, Morris M., and Evan J. Soltas. 2023. "A Welfare Analysis of Occupational Licensing in US States." *Review of Economic Studies* 90 (5): 2481–516.

Kleiner, Morris M., and Evgeny Vorotnikov. 2017. "Analyzing Occupational Licensing among the States." *Journal of Regulatory Economics* 52 (2): 132–58.

Kleiner, Morris M., and Ming Xu. 2025. "Occupational Licensing and Labor Market Fluidity." *Journal of Labor Economics* 43 (3): 937–83.

Larson, Bradley, Ziao Ju, Adam Kapor, and Chuan Yu. 2020. "The Effect of Occupational Licensing Stringency on the Teacher Quality Distribution." NBER Working Paper 28158.

Law, Marc T., and Sukkoo Kim. 2005. "Specialization and Regulation: The Rise of Professionals and the Emergence of Occupational Licensing Regulation." *Journal of Economic History* 65 (3): 723–56.

Law, Marc T., and Mindy S. Marks. 2017. "The Labor-Market Effects of Occupational Licensing Laws in Nursing." *Industrial Relations* 56 (4): 640–61.

Lazear, Edward P., and James R. Spletzer. 2012. "Hiring, Churn, and the Business Cycle." *American Economic Review* 102 (3): 575–79.

Markowitz, Sara, E. Kathleen Adams, Mary Jane Lewitt, and Anne L. Dunlop. 2017. "Competitive Effects of Scope of Practice Restrictions: Public Health or Public Harm?" *Journal of Health Economics* 55: 201–18.

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

Molloy, Raven, Christopher L. Smith, and Abigail K. Wozniak. 2014. "Declining Migration within the US: The Role of the Labor Market." NBER Working Paper 20065.

Molloy, Raven, Christopher L. Smith, and Abigail K. Wozniak. 2017. "Job Changing and the Decline in Long-Distance Migration in the United States." *Demography* 54 (2): 631–53.

Murphy, Brendan. 2021. "USMLE Step 2 CS Canceled: What It Means for Medical Students." American Medical Association, February 4. https://www.ama-assn.org/medical-students/usmle-step-1-2/usmle-step-2-cs-canceled-what-it-means-medical-students.

Norris, Conor, Shishir Shakya, Edward Timmons, and Edgar Orozco. 2024. *Nurse Practitioner Full Practice Authority and Access to Primary Care in North Carolina*. West Virginia University Knee Regulatory Research Center.

Posmanick, Benjamin, Bobby Chung, and Alexander Obie. 2024. "Licensed Professionals and Corporate Board Performance: The Effect of the Sarbanes-Oxley Act on the Audit Committee." University of South Florida Department of Economics Working Paper 2024–03.

Rebolledo, Nicolas Acevedo, Kathryn J. Blanchard, and Stephanie Riegg Cellini. 2025. "Cosmetology Gets a Trim: The Impact of Reducing Licensing Hours on Colleges and Students." NBER Working Paper 33936.

Smith, Adam. 2016. *The Wealth of Nations*. Aegitas.

Springman, Jen. 2025. *Workers Welcome: ND, WV Champion Universal License Recognition*. Goldwater Institute.

Sweetland, Kyle, and Dick M. Carpenter II. 2022. *Raising Barriers, Not Quality: Occupational Licensing Fails to Improve Services.* Institute for Justice.

Thornton, Robert J., and Edward J. Timmons. 2013. "Licensing One of the World's Oldest Professions: Massage." *Journal of Law and Economics* 56 (2): 371–88.

Thornton, Robert J., and Edward J. Timmons. 2015. "The De-licensing of Occupations in the United States." *BLS Monthly Labor Review* (May): 1–20.

US Department of the Treasury Office of Economic Policy, Council of Economic Advisers, and Department of Labor. 2015. *Occupational Licensing: A Framework for Policymakers.* White House.

Xia, Xing. 2021. "Barrier to Entry or Signal of Quality? The Effects of Occupational Licensing on Minority Dental Assistants." *Labour Economics* 71: 102207.

Journal of Economic Perspectives—Volume 40, Number 1—Winter 2026—Pages 191–214

Asian Immigration to the United States in Historical Perspective

Hannah M. Postel

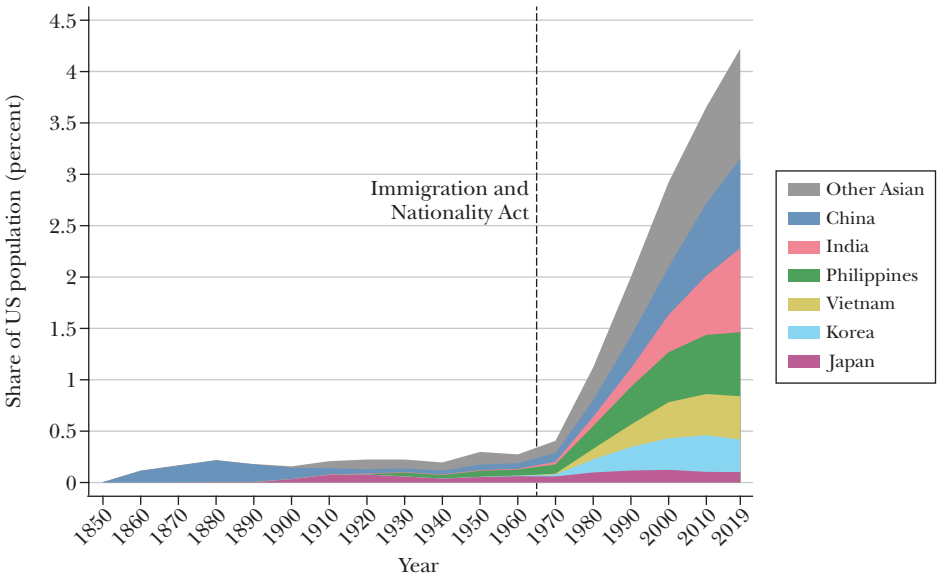
Asian Americans are the fastest-growing racial group in the United States, projected to become the country's largest immigrant group by 2050 (Budiman and Ruiz 2021). The Asian-identifying population is now nearly 8 percent of the US population, having more than doubled since 2000 (Tran 2024). This population growth is driven mainly by immigration, a trend unique to Asian Americans (Lee et al. 2024). Two in three Asian Americans are immigrants, while nine in ten are immigrants or children of immigrants (Tran 2024). In 2013, China and India overtook Mexico as the top sending countries of US-bound immigrants (Jensen 2015).

The trajectory of immigration from Asia has been distinctly shaped by US policy. A series of Asian exclusion laws beginning in 1875 severely restricted immigration from the region for nearly a century. As a result, the US Asian population remained under 1 percent of the total population until 1970. Only with the passage of the 1965 Immigration and Nationality Act were Asian immigrants able to enter the United States more freely. According to historian Roger Daniels (1990, p. 2), "immigrants from Asia were in a special category: they were the first to be singled out for discriminatory treatment by our immigration laws, and they were the only ethnic groups ever to be totally excluded." Outside of dedicated immigration policy, US foreign policy has also played a major role in shaping immigration flows through imperial annexations (the Philippines from 1898 to 1946) and international conflicts (Korea from 1950 to 1953 and Vietnam from the early 1950s to 1973).

■ *Hannah M. Postel is Assistant Professor of Public Policy, Sanford School of Public Policy, Duke University, Durham, North Carolina. Her email address is hannah.postel@duke.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251453.

Figure 1
Share of US Population Born Abroad by Major Asian Country



Source: Author's calculations from the 1850—2010 Census and 2019 American Community Survey.
Note: This figure plots the population share of major Asian immigrant groups from 1850 to 2019.

The effects of these changes can be seen in Figure 1, which plots the population shares of foreign-born Asian Americans from 1850 to 2019. The six largest Asian origin countries—China, Japan, India, the Philippines, Korea, and Vietnam—are shown individually, with the remaining origin countries aggregated as "other Asian." Populations for all six Asian national origins remained extremely low (below half a million in total) until the 1965 Immigration and Nationality Act abolished previous restrictions. In 1960, the entire Asia-born population could have resided in the 24th-largest US city: Atlanta. By 2019, the same population would fill the equivalent of two New York Cities, an increase of approximately 2,700 percent.

Beyond restricting immigration flows, US Supreme Court decisions—primarily in *Ozawa v. United States* (260 US 178 [1922]) and *United States v. Bhagat Singh Thind* (261 US 204 [1923])—also prevented Asian immigrants from naturalizing as US citizens until the 1940s. Indeed, Asian immigrants (particularly Chinese immigrants) were the progenitors of the entire US immigration system as we know it. Chinese exclusion was "the main catalyst that transformed the United States into a gatekeeping nation" (Lee 2003, p. 9). Now-widespread immigration enforcement measures such as visas and registration documents, entry interviews, detention, and deportation were originally designed to police Asian immigrants.

Despite its historical and contemporary resonance, Asian immigration has been relatively understudied by quantitative social scientists (Lee and Ramakrishnan 2021; Postel 2024).[1] This pattern is not new; in one seminal account of American nativism, Higham (1955) described Asian exclusion movements as merely "tangential." This paper synthesizes the limited existing research on historical Asian immigration to the United States and how this history has had lasting impacts on Asian immigrants' demographic and socioeconomic characteristics today. Given the centrality of US policy in shaping Asian immigration flows both past and present, I situate these analyses in the context of three major policy regimes.

"Asian immigration" encompasses a wide set of heterogeneous flows shaped by origin-country circumstances and nationality-specific US immigration policies. Most of this paper therefore focuses separately on six major origin countries: China, Japan, India, the Philippines, Korea, and Vietnam. In 2021, these countries accounted for 80 percent of the Asia-born population in the United States (Budiman and Ruiz 2021). When possible, the paper also highlights the experiences of smaller Asian origin groups. The analysis will emphasize the diversity of Asian immigrant experiences temporally, geographically, and across a range of socioeconomic outcomes. This paper focuses solely on first-generation, foreign-born immigrants, setting aside issues of intergenerational mobility and later-generation outcomes. However, the term "Asian American" is used to refer to the entire population of Asian origin, including US-born individuals of Asian descent, unless explicitly noted otherwise.

The paper first introduces the three main policy eras of interest—Asian Exclusion (1882–1943), Asian Restriction (1943–1965), and Growth and Diversification (1965–present)—with an eye to how each shaped the numbers and characteristics of Asian immigrants admitted. I then discuss geographic residence patterns of the population born in Asia, including residential segregation, and the educational, occupational, and broader economic outcomes of Asian immigrants.

Policy Eras and Immigrant Selection

The Asian Exclusion Era (1882–1943)

The tale of Asian exclusion is one of cyclical US labor needs and nativist backlash. The first Asian immigrants arrived in the United States in the mid-nineteenth century, and the first sizable Asian immigrant community was Chinese (Daniels 1990). At the outset, most Chinese immigrants were contract laborers hired for temporary work in mines or on major infrastructure projects (like the Transcontinental Railroad). These migrants were both searching for economic opportunity and escaping sociopolitical disorder at home (Chinn, Lai, and Choy 1969).

[1] There is of course an extensive, largely excellent historical literature built mainly on archival and oral histories. For comprehensive overviews of Asian immigration to the United States, see Lee (2015) and Takaki (1998). On Chinese immigrants, see Lew-Williams (2018, 2025); on imperialism and immigration, see Jacobson (2000); on Indian immigrants, see Bald (2013).

The Chinese became a target of American racial animus and the nation's first federal immigration laws during the late nineteenth century. By the 1870s, about 15,000 to 20,000 permanent Chinese immigrants were arriving each year (based on my own calculations from annual US Immigration and Naturalization Service Statistical Yearbooks, starting in 1850). For perspective, the total US population in 1870 was about 38 million, or one-ninth its current level. The 1875 Page Act excluded Chinese women, thereby precluding Chinese family formation for almost a century.[2] The 1882 Restriction Act limited future Chinese immigration to the United States and was extended in 1888 to "cut off circular Chinese migration, [pour] ten times the funding into enforcement, [eliminate] many avenues for fraud, and [develop] new systems of management and deportation" (Lew-Williams 2014, p. 27).

New Chinese immigration then dropped to near-zero for decades. The Chinese immigrant population in the United States peaked at 106,688 in 1890, falling to a low of 45,757 in 1920.[3] Besides a decrease in total numbers, another major change was the nearly order-of-magnitude decrease in the sex ratio from 1880 to 1940. Due to the earlier exclusion of Chinese women, the ratio of Chinese men to women was more than 20:1 from 1880 through 1910. However, by 1930 and 1940, the ratio had fallen to four or five to one. Due to selective immigration and emigration, the Chinese population in the United States remained largely working age during the entire exclusion period (Chew and Liu 2004; Chew, Leach, and Liu 2009).

When policies restricting Chinese immigration in the 1880s led to labor shortages, employers turned to Japanese immigrants looking for an escape from the sociopolitical upheaval of the Meiji Restoration (Takaki 1998). As the Chinese population fell into the early 1900s, the number of Japanese immigrants grew from only 2,292 in 1890 to over 80,000 by 1920.

However, this population in turn generated resentment, and Japanese immigration was also restricted by the 1907 "Gentlemen's Agreement." This was a bilateral agreement in which Japan's government agreed to limit how many passports it would issue to Japanese laborers seeking to emigrate to the United States. In turn, the US government agreed to end certain discriminatory practices against Japanese immigrants, like segregated schools in California. The agreement did allow wives, children, and parents to join family members who already resided in the United States. This had important demographic consequences: the number of Japanese-born women in the United States nearly tripled between 1910 and 1920. From a highly skewed sex ratio of over 20 men to every woman in 1900, by 1920 the sex ratio for the immigrant Japanese population was nearly balanced.

Though Asian exclusion policies originally focused on the Chinese and Japanese as individual groups, it was expanded to a continent-wide "barred zone" in

[2] Anti-miscegenation laws also precluded Chinese men from marrying white women (Pascoe 2009). The combination of these restrictions led to fewer than 10 percent of Chinese Americans being able to trace their ancestry back three generations (Lew-Williams 2018).

[3] In the paragraphs that follow, numbers about the size of immigration populations from various countries, as well as average age and male/female sex ratios, are all based on author's calculations according to US Census Bureau Data accessed via IPUMS USA.

1917. Since Chinese and Japanese were already restricted, and other Asian groups were not yet present in the United States, the main additional group affected was Indian immigrants. This group consisted of only a few thousand individuals, mainly laborers from the Punjab. The Indian-born immigrant population in the United States did not exceed 10,000 until after 1950.

The Asian exclusion movement culminated in the 1924 Johnson-Reed Act. This law superseded previous restrictions, imposing strict country quotas on immigration from all national origins. Because most Asian groups had already been restricted or excluded in some form, the main effect of the 1924 Act was to cap Japanese immigration, more strictly limiting the family reunification made possible under the Gentleman's Agreement.

These Asian immigration policy restrictions in the 1920s then interacted with a US economic downturn during the Great Depression of the 1930s. During this period Chinese immigration fell to under 5,000 annually and Japanese immigration dropped to under 2,000. Geopolitical upheavals—including World War II and China's Communist revolution—contributed to the persistence of these low numbers until 1965 (White, Biddlecom, and Guo 1993).

Only one Asian origin group during this period was granted freer entry to the United States. Following the end of the Spanish-American War in 1898, Spain ceded the Philippines to the United States for a payment of $20 million. Filipinos were thus considered US nationals at this time and unrestricted by the exclusionary policies described above. The Filipino immigrant population grew from essentially zero in 1900 to 45,208 by 1930. The largest period of growth came between 1920 and 1930, when immigrants of other national origins were restricted but Filipinos were not.

Hawaiian sugar-planters first recruited Filipino laborers in 1909 to replace Japanese workers, deliberately targeting illiterate men who were either single or willing to leave families behind. By 1931, about 113,000 Filipinos had migrated to Hawaii (which did not become a US state until 1959), with 19,000 subsequently moving to the West Coast to join more than 31,000 compatriots who had entered California directly during the 1920s (Liu, Ong, and Rosenstein 1991). However, Filipino migration was then curtailed in 1934 under the Tydings-McDuffie Act, which deemed Filipinos "foreign aliens" subject to a quota of 50 new immigrants per year. The overall population of Filipino immigrants therefore stopped growing, remaining at about 45,000 during this period. The Philippines became formally independent from the United States in 1946.

From Exclusion to Restriction (1943–1965)

A series of changes to immigration policies—largely tied to US involvement in World War II—loosened some prior barriers. Although quota numbers for new immigrants remained low (approximately 100 from each country per year), these new policies also allowed migration for certain demographic groups and enabled naturalization for others. Though total immigration numbers did not increase until after 1965, the laws passed during this era were central to Asian immigrant family formation and socioeconomic mobility (via the benefits afforded by citizenship).

Table 1
Major US Immigration Legislation Affecting Asian Origin Groups, 1943–1952

Year	Act	Affected group(s)	Naturalization	Immigration quota/pathway	Key provisions
1943	Magnuson	Chinese	Eligibility granted	105/year quota	Repealed Chinese Exclusion
1945	War Brides	All (spouses and children of US military)	If group eligible	Quota-exempt	Marriage to US service member required
1946	Luce-Celler	Indian, Filipino	Eligibility granted	100/year each	
1946	Alien Fiancées/ Fiancés	All (fiancé(e)s of US citizens)	If group eligible	Quota-exempt	Engagement to US citizen required
1952	McCarran-Walter	All Asian nationalities	Removed remaining racial bars (Japanese, Korean)	100/year per country	Established labor preference system

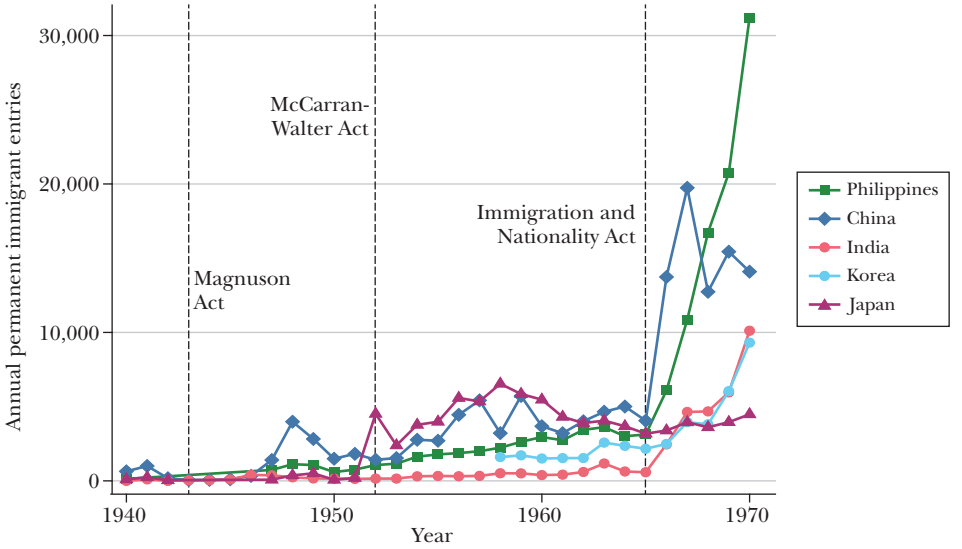
Table 1 below lists the main elements of the major immigration policies passed during this era.

Figure 2 below shows the gradual increase in yearly immigrant entries during this period. The Chinese immigrant population almost doubled from 1940 to 1950, but still remained under its 1890 peak of 106,000. Though the Indian population grew by about 50 percent after Luce-Celler's passage in 1946, it remained small at approximately 13,000.

Foreign-born persons engaged to be married to US citizens, along with the spouses or children of US military service members, were exempted from country quotas. Approximately 100,000 new Asian immigrants entered through these provisions; these new immigrants were also eligible for naturalization and jump-started Asian family formation in the United States. The War Brides Acts of 1945 had particularly significant effects on specific groups. From 1946 to 1965, another 33,000 Filipino immigrants entered the United States, with nearly half (16,000) arriving as wives of American servicemen (Liu, Ong, and Rosenstein 1991). Similarly, Japanese immigration in the 1950s consisted more of war brides than refugees, reflecting the ongoing American military presence in Asia (White, Biddlecom, and Guo 1993). In part due to these laws, Japanese immigrants were the largest Asian origin group in the United States in 1960, followed by Filipinos.

The repercussions of US overseas conflicts also facilitated the immigration of a new Asian origin group: Koreans. The first group of Korean immigrants had arrived in Hawaii in 1903, escaping famine and political turbulence at home—also, new workers were needed to fill labor gaps created by early legislation excluding Chinese and Japanese laborers. However, this flow was small and short-lived, as emigration was barred after Japan's 1910 annexation of Korea. Under the 1924 Immigration Act, Koreans were not allowed to immigrate to the United States until after the Korean War (1950–1953).

Figure 2
Yearly Permanent Immigrant Entries to the United States, 1940–1970



Source: Author's calculations from annual US Immigration and Naturalization Service Statistical Yearbooks, 1850–2022.
Note: This figure shows the slow growth of Asian immigration from major origin countries during the Restriction era (1943–1964). Entries for each group remained well under 10,000 annually.

By the 1950s, Korean immigration mainly consisted of war brides and post–World War II adoptions of Korean orphans. The Korean population in the United States grew tenfold between 1950 and 1960 following the Korean War, albeit remaining quite small at 11,171. Korean immigration patterns in the 1950s and 1960s reflected military connections more than refugee flows. Of the 17,000 Korean arrivals between 1950 and 1965, most were war brides. Only about 3,000 refugees arrived during the 1950s and 1,300 during the 1960s (White, Biddlecom, and Guo 1993). The Korean immigrant population did not grow extensively until after 1970, reaching a total of 289,885 in 1980.

Another crucial change during this era—sequentially extended to different Asian immigrant groups—was the ability to become a naturalized citizen. Before these laws, Asian immigrants remained permanently ineligible for citizenship regardless of length of residence. This new eligibility to citizenship enabled a range of economic and civic benefits, including the ability to own property, access occupations restricted to citizens, and vote.[4]

[4]Most states had enacted "alien land laws" prohibiting property ownership by "aliens ineligible for citizenship"; the federal naturalization provision removed these state-level barriers. For more on how state- and local-level policy restrictions affected Asian immigrants, including civic and occupational restrictions, see Lew-Williams (2025).

Finally, the 1952 McCarran-Walter Act introduced labor preference priorities and created several new visa categories that would become central to twentieth- and twenty-first-century Asian immigration. Immigrants with "special skills" were considered exempt from country quotas, thereby shifting the socioeconomic background of incoming immigrants. The H1 and H2 visas were introduced for temporary high- and low-skilled immigrants, respectively.

Growth and Diversification (1965–present)

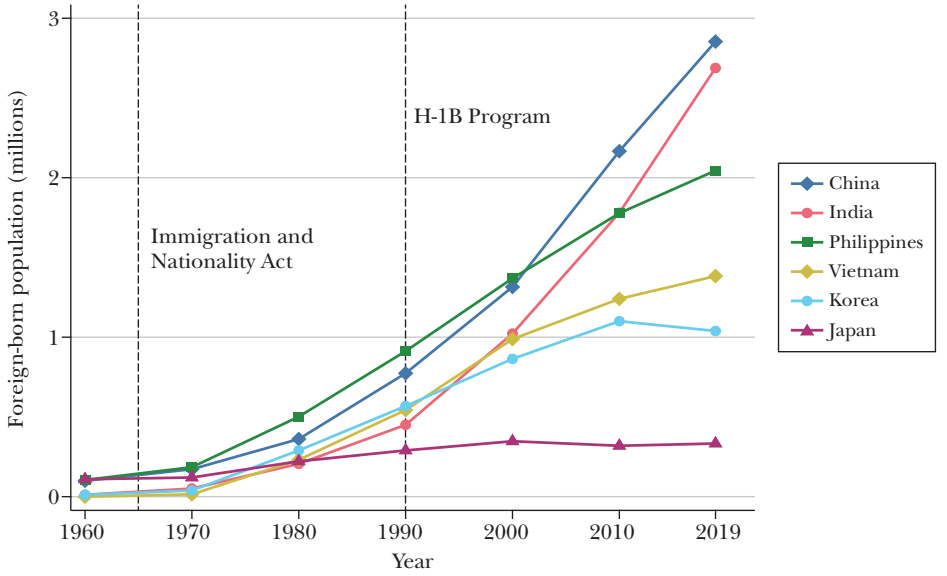
Asian immigration to the United States expanded and diversified dramatically beginning in the 1960s. Three major changes drove this shift. First, after over a century of restriction and at times full exclusion of Asian immigration, national quotas for immigration were removed under the Hart-Celler Act of 1965. Second, this law also diversified visa channels and opened new pathways for entry, especially for family members and highly skilled professionals. Third, refugee programs admitted large numbers of Asians fleeing conflicts in their birth countries, beginning with the Indochina Migration and Refugee Assistance Act of 1975 and later institutionalized under the Refugee Act of 1980.

The result of these changes was extraordinary population growth for Asian migration, as Figure 3 illustrates. The overall Asian population in the United States has increased more than 20-fold since 1960. China, Korea, the Philippines, and "other Asian" groups followed similar trajectories, while Indian and Vietnamese population growth was even larger. The Japanese immigrant population grew much more modestly than the other groups.

The emphasis of the Hart-Celler Act of 1965 on family reunification and professional skills also shaped the composition of new arrivals. Many of the first post-1965 Asian immigrants were highly educated and professionally skilled, including doctors, nurses, and engineers (Lee et al. 2024). Over the longer term, the Hart-Celler framework established multiple permanent entry channels that structured Asian immigration for decades to come. Its legacy is visible in the contemporary distribution of entry categories: in 2022, 41.9 percent of Asian immigrants arrived under family sponsorship, 44.8 percent under employer sponsorship, 5.8 percent as refugees or asylees, and 4.3 percent on a diversity visa (Tran 2024).

Refugee admissions became a defining pathway of Asian immigration in the late 1970s and early 1980s. After the end of US involvement in the Vietnam War in 1973, the Indochina Migration and Refugee Assistance Act of 1975 marked the beginning of large-scale Vietnamese migration to the United States, which—like earlier Korean migration—began in force following US involvement in an overseas conflict. The 1980 Refugee Act institutionalized refugee admissions and expanded annual ceilings. Within five years of the first evacuations, the Vietnamese population in the United States had grown to more than 250,000, with refugees resettled across the country. A second wave of "boat people" arrived between 1978 and the mid-1980s. Over time, Vietnamese migration has shifted: most now arrive through family-based channels, with 53 percent admitted through family sponsorship and 44 percent as immediate relatives of US citizens.

Figure 3
Foreign-Born Population in the United States from Major Asian Countries



Source: 1960–2010 US Census, 2019 American Community Survey.
Note: This figure shows the massive growth of major Asian immigrant groups following the 1965 Hart-Celler Act. The total Asian immigrant population grew more than 20-fold from 1960 to 2019.

The combination of these changes dramatically diversified the regional origins of Asian immigration. In 1960, 80 percent of the US Asian-born population was East Asian (from China, Japan, and Korea). Large-scale refugee resettlement in the 1970s increased the Southeast Asian share, especially from Vietnam and Cambodia, and since 1980 South Asians (mainly from India) have dominated the flow. By 2020, East Asia, Southeast Asia, and South Asia each accounted for about one-third of the Asian-born population (Hanna and Batalova 2021), creating what Lee and Ramakrishnan (2021, p. 4) call "unprecedented diversity within the US Asian population."

More recently, high-skill temporary migration has become an important pathway. Since the early 1990s, most new high-skill migrants from Asia have entered on temporary work or student visas (as Khanna discusses in this issue). The large-population countries of China and India have been the largest adopters, each receiving more than three million temporary visas between 1997 and 2023. By comparison, Koreans received 1.1 million such visas, Japanese 798,000, Filipinos 214,000, and Vietnamese 250,000 (US Department of State 2025).

Another notable trend is the rise of undocumented Asian immigration: by 2019, one in six Asian immigrants was undocumented (Warren 2021). That year, there were about 390,000 unauthorized Indian nationals and 330,000 unauthorized Chinese nationals living in the United States (INS 2022). In both 2023 and 2024, Indian- and Chinese-born individuals together accounted for around 3 percent of

US Customs and Border Patrol encounters at the US Southern border (approximately 60,000 encounters for each nationality).[5]

Taken together, the three policy eras show how fundamentally US law and foreign policy have shaped Asian immigration. Exclusion in the late nineteenth and early twentieth centuries restricted or eliminated entry for most Asian groups, while US imperial expansion created selective exceptions, most notably for Filipinos. Mid-century reforms during and after World War II began to legalize naturalization and open small but significant pathways for family formation. The watershed came with the Hart-Celler Act of 1965, which abolished entry quotas and diversified visa channels, followed by large-scale refugee admissions from conflicts in Korea and Vietnam. These changes have produced a larger and more diverse first-generation Asian immigrant population with important implications for residential and economic outcomes.

Geographic Settlement Patterns and Residential Outcomes

Asian immigrants have historically concentrated on the West Coast, initially because of proximity to Pacific ports and later through the growth of these established communities. In 1960, two-thirds of Asian immigrants lived in California and Hawaii; today, nearly one-half still reside in the US West region, with about one-third in California alone. But over time, new regions have grown in importance: by 2021, 24 percent of Asian immigrants lived in the South, 19 percent in the Northeast, and 12 percent in the Midwest (Budiman and Ruiz 2021). The continued growth and diversification of Asian immigration flows have contributed to this increased geographic spread.

These geographic shifts matter because the places in which immigrants settle shape their "contexts of reception": the local labor markets, housing conditions, institutions, and social environments that structure opportunity and assimilation (Portes and Rumbaut 2006). For example, regional differences in job availability, cost of living, and school quality will influence both residential choices as well as economic outcomes and long-run mobility (Waldinger 1996; Alba and Nee 2003). Settlement geography also affects the degree of contact with co-ethnic communities and the likelihood of encountering discrimination or segregated neighborhoods (Zhou 1992; Frey 2022). As Asian immigrants have dispersed beyond their historical centers of settlement, they have encountered a wider set of local conditions shaping their socioeconomic trajectories.

Geographic Dispersal Patterns

In general, older groups such as Chinese, Japanese, and Filipinos remain concentrated in the same areas where the first immigrants arrived, while newer

[5]An "encounter" is any interaction US Customs and Border Protection agents have with a noncitizen attempting to enter the United States, including those who are turned away at a port of entry for being inadmissible, caught crossing between ports of entry, or who present themselves at a port of entry to claim asylum.

groups like Koreans and especially Indians are more dispersed. These trends are visible in state-level maps by origin group in Figure 4, which show the enduring dominance of California as a settlement hub alongside the gradual spread of Asian populations into the Northeast, Midwest, and especially the South after 2000.

The original geographic patterning of Asian American communities was shaped by each group's policy era of entry. For example, many early Korean immigrants arrived as spouses of US military members; this led to more living outside of traditional ethnic enclaves (White, Biddlecom, and Guo 1993). Japanese internment during World War II—and the ensuing government-encouraged resettlement—scattered families across the Midwest and East (Grossman, Khalil, and Panza 2025; Hayashi 2021). Vietnamese refugees arriving in the 1970s and 1980s were deliberately dispersed nationwide (Zhou and Bankston 1998).

Entry pathways continue to shape geographic residence. Recent refugee-origin groups such as Hmong, Cambodians, Laotians, and Bhutanese have been resettled across the Midwest and South, including Minnesota, Wisconsin, Ohio, and Pennsylvania (Budiman and Ruiz 2021). By contrast, Bangladeshi and Pakistani immigrants, arriving mainly through employment and family channels, have concentrated in major metropolitan areas like New York, Chicago, Houston, and Washington, DC (Taylor and Ruiz 2017).

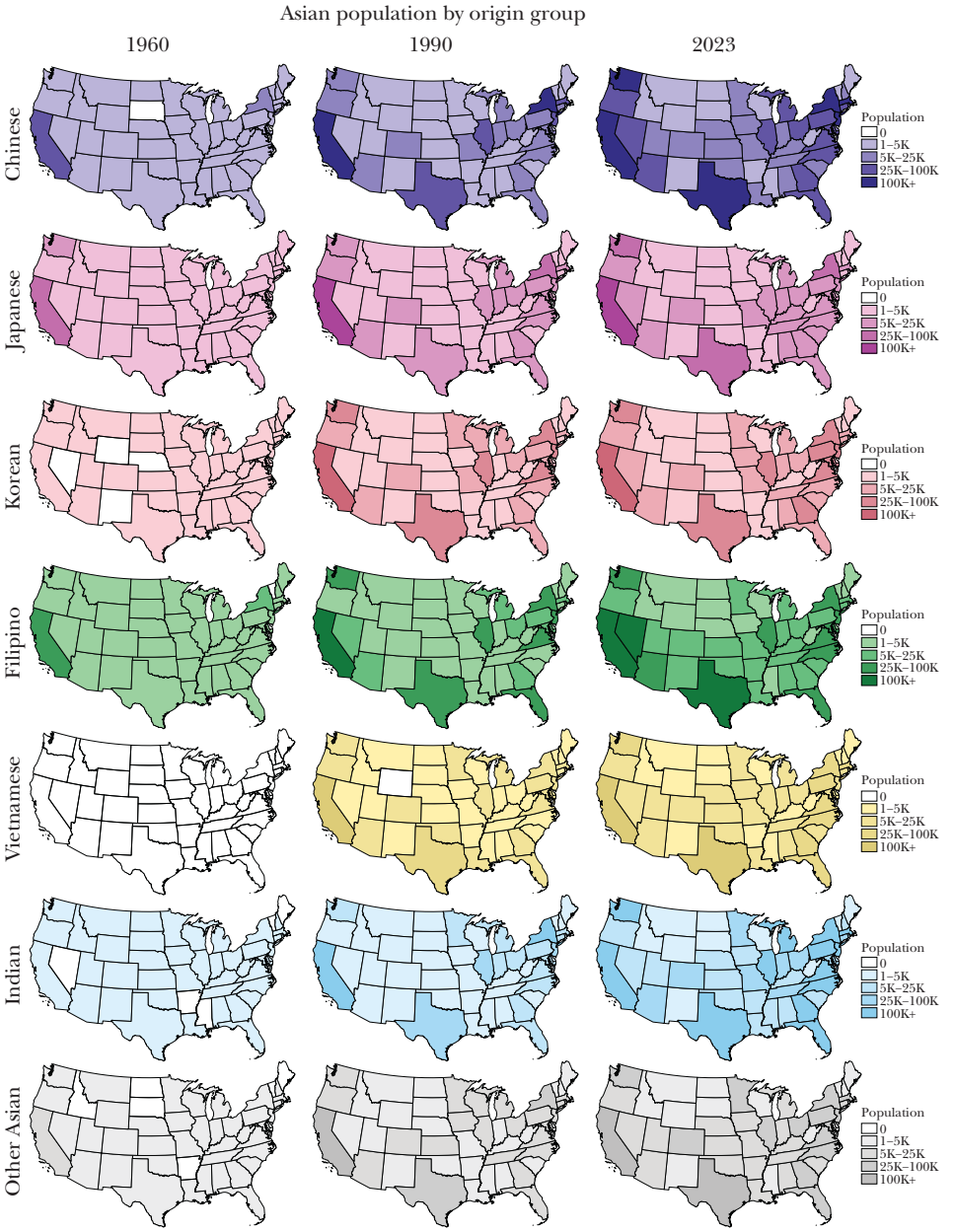
The Asian immigrant community has become increasingly geographically dispersed since 2000, with growing concentrations across the Midwest and South. Filipino and Japanese immigrants remain most concentrated on the West Coast, while Chinese and Koreans now show a strong bicoastal presence in the Northeast. Indians are the least concentrated, with roughly equal shares across US regions. Certain counties across the Mountain West, Plains, and Deep South that had negligible Asian populations in 1960 now show visible communities.

The most rapid growth in Asian populations between 2010 and 2020 occurred in so-called "New Immigrant Destinations," or areas with historically small immigrant populations recently experiencing rapid inflows. Starting from an admittedly low base, the Filipino population in North Dakota grew by 186 percent, the Indian population in North Carolina by 104 percent, and the Vietnamese population in Montana by 112 percent. Alabama, Idaho, Utah, and Wyoming have also experienced sizeable growth in their Asian American populations (AAPI Data 2024). These patterns mirror broader immigrant residence trends shaped by regional economic restructuring and secondary migration from high-cost locations (Singer 2004; Flippen and Farrell-Bryan 2021).

At a more granular level, the concentration of Asian immigrants in major urban areas has shifted. One way to illustrate this evolution is to look at the concentration of Asian immigrant groups in the three largest "gateway" metropolitan areas over time, as shown in Figure 5. The term "gateway" refers to metropolitan areas or regions that have historically served as initial settlement destinations for newly arrived Asian immigrants: Los Angeles, New York, and San Francisco.

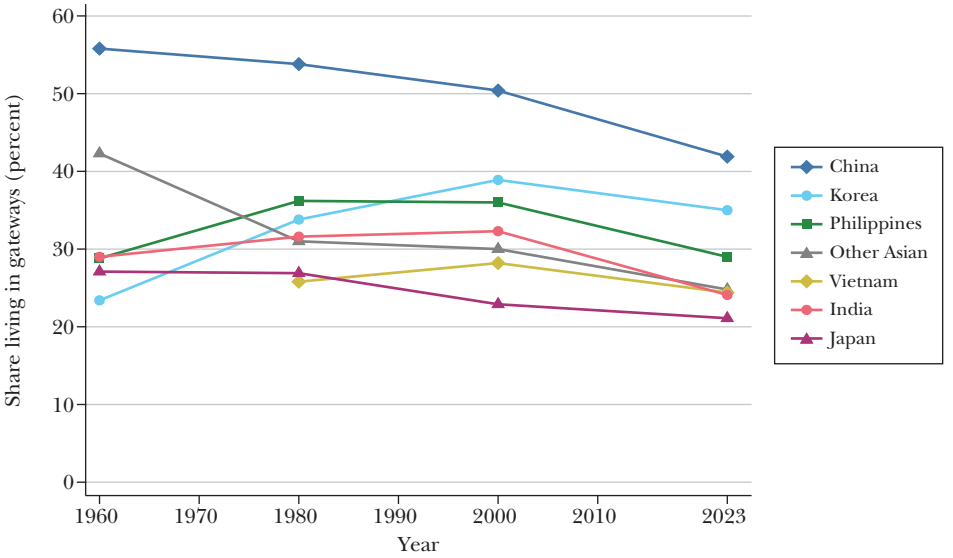
Chinese immigrants have been uniquely clustered across time; in 1960, fully 56 percent lived in the three major metros. "Other Asian" groups were also heavily

Figure 4
State-Level Asian Population by Country of Origin



Source: 1960 and 1990 US Census, 2023 American Community Survey.
Note: These maps plot the dispersion of major Asian immigrant communities across the United States, showing both increased density and spread from 1960 to 2023.

Figure 5
Asian Immigrant Residence Concentration in Gateways, 1960–2023



Source: US Census 1960–2000, American Community Survey 2023. The primary gateways for Asian migration are Los Angeles, New York, and San Francisco, which are measured here as consistent 2020 Metropolitan Statistical Areas.
Note: This figure shows Asian immigrant concentration in major "gateway" communities from 1960–2023.

concentrated in 1960 (over 40 percent in gateways). For Filipinos, Indians, Japanese, and Koreans, the share was closer to one-quarter. Even these levels represented much stronger concentration than the US population as a whole: only 14.3 percent of all Americans lived in those same three metros in 1960. Over the following decades, the share of most groups in the three "gateway" metros declined or stayed constant. Only Koreans became more notably more concentrated in gateways over time. By 2023, Chinese and Korean immigrants had the highest population share living in gateways.

Settlement patterns have also shifted *within* metropolitan areas. Since the 1980s, Asian Americans have become the most suburbanized minority racial group: by 2020, nearly two-thirds lived in suburbs (Frey 2022). This has been driven in part by the emergence of "ethnoburbs," or suburban middle-class ethnic neighborhoods (Li 1998). Of the 1,863 census tracts in the 50 largest metropolitan areas identified as ethnoburbs, 86 percent experienced Asian American population growth between 2010 and 2020 (Kye 2023). Thus, while traditional gateways remain dominant for Asian immigrants, dispersal to new regions and increased suburbanization have shifted residential patterning.

Residential Segregation Patterns

Another important residential outcome is segregation, or the degree to which different groups live together or apart. In general, high segregation means different

groups concentrate in separate spaces; low segregation indicates more mixed residential patterns.

Historical Asian segregation patterns remain poorly understood. At the turn of the twentieth century, Chinese immigrants in San Francisco exhibited unusually high segregation (Eriksson and Ward 2019). Yet despite Chinatown's narrative prominence, most Chinese immigrants at the time lived in less concentrated rural areas across California (Postel 2025). Over the twentieth century, as population numbers fell, segregation for Chinese and Japanese immigrants also declined, reaching levels comparable to European immigrants (Eriksson and Ward 2019).

Today, people of Asian origin experience the lowest segregation from whites of any minority group (Hwang and McDaniel 2022). But this broad category hides important variation. In fact, "every Asian nationality except Japanese is more segregated from whites than are Asians as a broad category" (Logan and Zhang 2013, p. 1). Intra-Asian segregation is also significant: different Asian groups are as segregated from one another as they are from the white population (Kim and White 2010). Asian residential patterns vary significantly across different types of metropolitan areas, with less clustering in nontraditional destinations (Lee 2021).

These patterns should be understood not only as reflecting potential preferences of Asian groups for co-ethnic living, but also as a reflection of white preferences to *not* live in majority-Asian areas. The increasing Asian concentration in ethnoburbs was often complemented by white exit: in over half of the tracts where Asian populations grew between 2010 and 2020, white populations declined (Kye 2023; Boustan, Cai, and Tseng 2024).

It is also important to remember that while historically segregated neighborhoods were often disadvantaged, that is not necessarily the case today. The modern "ethnoburbs" mentioned above are characterized by both a high concentration of Asian immigrants as well as relatively high income. As a result, Asian neighborhoods are often comparable—or superior—to white neighborhoods on socioeconomic measures. Unlike African Americans and Hispanics, whose residential segregation today continues to imply lower neighborhood quality, Asian settlement patterns could be described as "separate but equal (or even more than equal)" (Logan and Zhang 2013, p. 14).

Occupational Patterns and Economic Outcomes

Asian immigrant socioeconomic patterns have evolved dramatically over time, shaped by both US labor demands and the policy regimes that determined who could enter and under what circumstances. Apart from determining which national origin groups could enter, many of the laws discussed earlier also specified credential requirements and limited which occupations new immigrants could access. Immigration policies shaped not only the size of Asian immigrant populations but also their socioeconomic trajectories.

A central theme across these policy eras is the shifting relationship between the skills Asian immigrants brought with them and the work they were able to perform once in the United States. Two mechanisms shaped this relationship. First, immigration law determined who could enter. Before 1965, skill-based admissions existed through narrow, racially circumscribed exceptions that were not designed around US labor-market needs (Ngai 2004; Lew-Williams 2014). The 1965 Hart-Celler Act introduced large-scale, race-neutral employment preferences that made education and occupational training a systematic basis for entry (Daniels 1990; Lee 2015). Second, domestic labor-market structures determined how immigrants' skills translated into jobs: during the Exclusion and Restriction eras, racialized barriers sharply limited access to skilled occupations and funneled even high-skilled migrants into restricted niches (Chan 1984; Takaki 1998). After 1965, credentialed professionals more often found work aligned with their training, although humanitarian admissions continued to produce occupational downgrading (Lee and Zhou 2015).

Educational Achievement

The wave of highly educated arrivals following the 1965 Hart-Celler Act are often described as "hyperselected"; that is, they are more likely to have a college degree than both nonmigrants in their countries of origin and US-born individuals (Lee and Zhou 2015). For example, Chinese immigrants are 18 times more likely than nonmigrants in China to hold a college degree (53 percent versus 4 percent) and significantly more likely than US-born non-Hispanic whites to do so (33 percent) (Tran et al. 2018). This group has received a disproportionate amount of scholarly attention, particularly as regards their exceptional educational outcomes.[6]

However, some groups of Asian immigrants showed high educational attainment even during the Asian exclusion period. According to the author's calculations from US Census Bureau Data accessed via IPUMS USA, in 1940, 36 percent of Indian immigrants and 33 percent of Korean immigrants had attended college. Most other Asian immigrants held less than a high school education. Educational attainment of most first-generation immigrants declined between 1940 and 1950 (for example, Koreans declined to 20 percent college), whereas the share of Chinese immigrants who had attended college increased from 10 to 15 percent.

Following the 1965 Hart-Celler Act, educational trajectories diverged. India, China, and Korea exhibited large gains in the share of the immigrant population with bachelor's degrees or higher: India from 67 percent in 1990 to 78 percent in 2019, China from 47 percent to 63 percent, and Korea from 37 percent to 65 percent. Korean immigrants saw the fastest transformation, with a large drop in the share of those with a high school education or less (52 percent to 19 percent). Perhaps surprisingly given the entry channels mentioned above, the educational achievement of Filipino immigrants has remained relatively constant over time (around half with bachelor's degrees or more). Vietnamese immigrants continued

[6]For two exceptions to this trend, see Kim and Kim (2024) and Kim (2025). Both study Asian educational achievement prior to 1965.

to exhibit the lowest education levels of all groups, despite a slight increase from 19 percent college educated or more in 1990 to 32 percent in 2019.

Further divergence is seen at a more granular level. In 2019, India and China had large graduate-degree shares (51 and 44 percent, respectively), Korea and Japan were more balanced (roughly 45 percent bachelor's and 30 percent graduate), the Philippines remained concentrated in bachelor's degrees (61 percent), and Vietnam had the highest share with a high school education or less (58 percent). These credential distributions provide context for the occupational patterns in the following section.

Occupational Patterns

Early Asian immigration in the 1850s concentrated in manual labor sectors, with Chinese immigrants recruited as contract laborers for mining and major infrastructure projects like the Transcontinental Railroad. With the completion of the railroad in 1869, thousands of Chinese workers returned to California, where most sought work "as nonagricultural common labor" in manual and service occupations (Chan 1984, p. 290). Exclusionary legislation then imposed skill-based restrictions that reshaped occupational trajectories. The 1882 Chinese Exclusion Act barred Chinese laborers while allowing only exempt classes—for example, merchants, students, and diplomats—to enter. The 1917 Immigration Act extended this logic by including occupational exemptions for the "barred zone" that restricted most immigration from the entire Asian continent.

These regulations shaped durable niches. Under the earlier Chinese exclusion acts, certain groups of Chinese immigrants, like merchants and scientists, were still allowed to enter and remain in the United States. A 1915 court decision—*United States v. Lee Chee* (224 F. 447 [1915])—modified the entry exemption for Chinese merchants to include restaurant owners. In the following years, the number of Chinese restaurants in New York City nearly quadrupled and more than doubled again between 1920 and 1930 (Lee 2014). Hundreds of state and local licensing and workplace exclusion laws further channeled Chinese immigrants into—and out of—specific industries (Postel and Lew-Williams 2025). For example, Chin and Ormonde (2018) discuss rules related to fishing—which affected not only Chinese but also Japanese, Filipino, and Asian Indian sailors—while Chin and Chin (2022) discuss policies related to restaurants. By 1920, 58 percent of Chinese immigrants were employed in services—mostly restaurants and laundries—compared with 5 percent of native-born whites and 10 percent of foreign-born whites (Takaki 1998, p. 240).

Table 2 below shows the continued importance of eating and drinking establishments for Chinese employment. In contrast, high-skill Chinese arrivals since 1965 have been concentrated in industries including information technology (see Khanna in this issue).

Like the Chinese, Japanese workers were initially employed as laborers for railroads, mines, and lumber mills. In fact, they were hired to fill the exact jobs vacated under Chinese Exclusion. By the early 1900s, the Japanese had expanded into

Table 2
Top Employment Industry by Country of Birth, 1990–2019

	1990	2000	2010	2019
China				
Top industry	Eating/drinking	Eating/drinking	Eating/drinking	Eating/drinking
Percent labor force	17.5%	12.5%	13.9%	10.8%
India				
Top industry	Hospitals	IT	IT	IT
Percent labor force	9.6%	9.8%	13.0%	17.3%
Japan				
Top industry	Eating/drinking	Eating/drinking	Eating/drinking	Eating/drinking
Percent labor force	8.1%	6.7%	5.9%	6.2%
Korea				
Top industry	Eating/drinking	Eating/drinking	Eating/drinking	Eating/drinking
Percent labor force	9.1%	8.6%	9.8%	6.3%
Philippines				
Top industry	Hospitals	Hospitals	Hospitals	Hospitals
Percent labor force	16.3%	14.9%	16.9%	15.4%
Vietnam				
Top industry	Eating/drinking	Electrical machinery	Misc. personal services	Misc. personal services
Percent labor force	8.1%	8.1%	13.8%	16.6%

Source: US Census 1990–2010, American Community Survey 2019.
Note: The IT category refers specifically to "computer and data processing"; abbreviated for the sake of presentation. In the data for Vietnam, the overwhelming majority (86 percent) of individuals employed in "miscellaneous personal services" were employed as "personal appearance workers," likely in nail salons.

agriculture, despite being barred from owning land in many states under a series of Alien Land Laws. However, internment during World War II forced Japanese Americans to abandon farms and businesses (Grossman, Khalil, and Panza 2025). Most agricultural holdings were sold at depressed prices, and even after the war, returning to farming was difficult as many Alien Land Laws were still in place. By 1950, many Japanese Americans had entered urban wage work instead of returning to agriculture (Lin and Peri 2025; Takaki 1998). The occupational disruption of internment, compounded by continuing legal barriers, reshaped Japanese American economic trajectories in the postwar decades. Perhaps counterintuitively, Arellano-Bover (2022) finds that internment had a long-run positive effect on internees' annual income. He suggests two possible channels for this finding: changing location and occupations after internment, and information exchange facilitated by diverse networks in the camps.

Early Korean migration in the mid-twentieth century was a dual flow comprised of both agricultural laborers and higher-wage professionals. Early arrivals included clergymen, scholars, soldiers, students, and farmers; literacy rates were relatively high. Koreans worked widely as farm laborers, but also developed entrepreneurship

in hotels, restaurants, and retail. By 1980, 25 percent of Koreans were professionals or managers, and 40 percent of Korean household heads in Los Angeles County were self-employed (Takaki 1998). Korean immigrants continue to exhibit much higher self-employment rates than other Asian immigrant groups (Fernandez and Kim 1998).

Indian immigrants show the most striking occupational transformation. Early Punjabi Sikh arrivals concentrated in California agriculture; in 1940, 65 percent of Indian-origin residents worked in agriculture, and only 4 percent were professionals (Takaki 1998, p. 314). During the 1960s and 1970s, India's economy could not keep up with the number of college graduates it was producing, with 100,000 engineers out of work in India in 1974, and educated professionals sought opportunities abroad just as the 1965 Immigration Act offered visas to those with technical skills (Lee 2015, p. 296). Between 1966 and 1973, 83 percent of Indian immigrants entering under occupational categories had been professionals and technical workers in India (Takaki 1998, p. 446). The early waves of Indian migrants participated in the hardware boom of the 1970s and 1980s, and these migrants facilitated the hiring of Indian nationals during the 1990s software boom (Saxenian 1999).

Early Filipino immigrants were mainly employed in agriculture. This was a direct substitution of Japanese and Chinese workers who had been recently excluded. They also experienced similar student exemptions: the 1903 Pensionado Act brought elite students to US universities. The 1934 Tydings-McDuffie Act that restricted Filipino immigration aligned with lower labor demand during the Great Depression (although one interesting exception is the Filipino immigrants who joined the US Navy during World War II, as discussed in Lee 2015). It was not until 1946 that the Education Exchange Act opened migration channels for Filipino nurses; this program was expanded under the 1948 Exchange Visitor Program, seeding what became a long-term concentration in healthcare (Choy 2003). The US demand for healthcare professionals prompted the expansion of visas for medical workers in 2000, resulting in increased migration of nurses from the Philippines (Abarcar and Theoharides 2024). As Table 2 shows, hospitals have been consistently the largest industry of Filipino employment.

For Vietnamese immigrants, occupational trajectories were heavily influenced by the refugee resettlement process. The Indochina Migration and Refugee Assistance Act of 1975 and the Refugee Act of 1980 relied on state-sponsored resettlement agencies to help refugees find jobs in cooperating industries. As one example, Gulf Coast seafood companies hired and trained Vietnamese refugees as fishermen, oyster shuckers, and crab pickers (Starr 1981). This labor pipeline meant that many Vietnamese experienced occupational downgrading after arrival in the United States. A 1978 study found that while 40 percent of refugee household heads had been professionals in Vietnam, only 7 percent were professionals within two years of US entry (Takaki 1998, p. 454). Over time, Vietnamese immigrants have developed niches in personal services; beginning in the 1990s, nail salons have become a major industry, now employing nearly one-fifth of Vietnamese immigrant workers (Bankston and Zhou 2021; see also Table 2 above).

Table 2 presents some of the occupational patterns just discussed as well as the industrial concentration of each national origin group. It also shows the industrial concentration of each national origin group. Some groups (Vietnamese, Filipino, Indian) are much more concentrated in their top industry than others (Chinese, Japanese, Korean). These contemporary outcomes demonstrate both the lasting influence of historical entry channels and adaptation to new US labor demands.

Economic Outcomes and Inequality

The occupational selectivity created by Asian exclusion has had lasting implications for socioeconomic outcomes. Chinese immigrants were highly occupationally selected for nearly a century beginning in 1882. Two studies have shown that Chinese socioeconomic outcomes worsened accordingly (Chen 2015; Chen and Xie 2024). However, the human capital of Chinese immigrants' children increased; this may suggest an active response to discrimination through investment in the next generation (Kim 2025).

Educational advantage has not guaranteed proportional labor market success. While Asian Americans have the highest education levels of any major US racial group, earnings and mobility vary sharply by gender, nativity, national origin, and place of education (Lee et al. 2024). This dynamic has also been observed in historical context. For example, although Chinese and Japanese Americans achieved high educational mobility during the Asian exclusion period, this did not translate into occupational mobility at that time (Tian, Song, and Xie 2023).

Since 1970, median Asian household income has exceeded that of whites (Kochhar and Cilluffo 2018).[7] However, this overarching narrative masks considerable inter-group inequality. Indians are the most advantaged Asian immigrant group, with high educational attainment and low poverty. Vietnamese are least advantaged, with half the rate of college completion and double the poverty rate as Indians (Abramitzky et al. 2021; Tran 2024).

Intra-group inequality is also significant. In particular, Chinese immigration has been bimodal, with highly educated professionals at one end and low-wage workers at the other. Chinese immigrants therefore have extremely high income disparities, with a 90/10 ratio (that is, the ratio of income at the 90th percentile to income at the 10th percentile) of almost 19 (Tran 2024). Koreans also exhibited very high inequality (90/10 of 18.3) compared to the overall US ratio of 12.6 in 2018 (Horowitz, Igielnik, and Kochhar 2020). Other Asian immigrants were less unequal than the US average, particularly Japanese (10.9), Indians (9.6), and Filipinos (7.4) (Tran 2024).

In sum, Asian American socioeconomic attainment is best described as "a high average and a large dispersion" (Zeng and Xie 2004, p. 1076). Policy-driven occupational selection led to educational "hyper-selectivity" for many groups.

[7] This may be in part due to household structure, as Asian households are more likely multigenerational with more earners: 34 percent among Vietnamese immigrants and 25 percent among Chinese immigrants (Budiman and Ruiz 2021).

However, outcomes diverge sharply both within and across groups. These dynamics underscore how immigration law and labor demand have continued to shape occupational and economic outcomes across time.

Conclusion

Asian immigrants have been distinctive in US history as the first immigrant group defined as permanently "excludable," denied naturalization, and subjected to racialized restrictions long before the 1924 national origins quotas. Though other immigrant groups faced discrimination from native-born white Americans, most were considered legally white and could jumpstart their own socioeconomic assimilation through naturalization (Catron 2019).

Yet scholarship on Asian Americans remains limited, constraining our ability to understand how these historical processes shaped outcomes across generations. Over the past two decades, the top journals in five disciplines published on average only one article per year on Asian American topics. Data infrastructure continues to privilege the largest national-origin groups, making it difficult to study smaller and more vulnerable communities (Tran 2024). Aggregated statistics obscure the fact that differences within the Asian-origin population are often larger than differences between Asians and other ethno-racial groups (Tran 2024). These gaps also shape curricula and funding priorities, reinforcing a cycle in which Asian American experiences remain underrepresented (Lee and Huang 2021).

There is currently no existing panel study focusing on behavioral and social trends for Asian Americans (Tran 2024, p. 14). In addition, creating panel data by linking cross-sectional Asian immigrant records (for example, via Census data) is significantly harder than for other populations (Postel 2023). Without repeated observations, researchers cannot examine life-course trajectories, disentangle causal mechanisms, or compare Asian Americans directly with the well-documented experiences of Black, white, and Hispanic Americans (Lee et al. 2024).

The lack of panel data is particularly salient for immigrant communities, as cross-sectional data are often confounded. Cohort differences and selective return migration make it difficult to identify the processes underlying socioeconomic change. Work on historical European immigrants illustrates the potential stakes: once panel data became available, they overturned prevailing wisdom from cross-sectional studies. Abramitzky, Boustan, and Eriksson (2014) showed that European immigrants held similar occupations to natives and experienced similar rates of advancement, in contrast to the existing hypothesis that immigrants faced an earnings penalty upon arrival and converged with natives over time. Without similar data for Asian Americans, our understanding of their trajectories remains incomplete.

As Lee et al. (2024, p. 14) argued, "we cannot understand race, which is so central to the American social structure . . . if we do not incorporate the study of Asian Americans in our funding priorities and our research." Only by addressing

these research gaps can we fully understand how the US policy continues to shape Asian immigrant outcomes today.

■ *Thanks to Gaurav Khanna for formative discussions; Jeffrey Kling, Jonathan Parker, Sebastian Saling, Timothy Taylor, and Heidi Williams for helpful comments; and Braden Clark for excellent research assistance.*

References

AAPI Data. 2024. *AANHPI Community Counts Dashboard*. http://counts.aapidata.com/census2020/ (accessed on August 15, 2025).

Abarcar, Paolo, and Caroline Theoharides. 2024. "Medical Worker Migration and Origin-Country Human Capital: Evidence from US Visa Policy." *Review of Economics and Statistics* 106 (1): 20–35.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2014. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy* 122 (3): 467–506.

Abramitzky, Ran, Leah Boustan, Elisa Jácome, and Santiago Pérez. 2021. "Intergenerational Mobility of Immigrants in the United States over Two Centuries." *American Economic Review* 111 (2): 580–608.

Alba, Richard, and Victor Nee. 2003. *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Harvard University Press.

Arellano-Bover, Jaime. 2022. "Displacement, Diversity, and Mobility: Career Impacts of Japanese American Internment." *Journal of Economic History* 82 (1): 126–74.

Bald, Vivek. 2013. *Bengali Harlem and the Lost Histories of South Asian America*. Harvard University Press.

Bankston, Carl L., and Min Zhou. 2021. "Involuntary Migration, Context of Reception, and Social Mobility: The Case of Vietnamese Refugee Resettlement in the United States." *Journal of Ethnic and Migration Studies* 47 (21): 4797–816.

Boustan, Leah, Christine Cai, and Tammy Tseng. 2024. "JUE Insight: White Flight from Asian Immigration: Evidence from California Public Schools." *Journal of Urban Economics* 141: 103541.

Budiman, Abby, and Neil G. Ruiz. 2021. "Key facts about Asian origin groups in the U.S." *Pew Research Center* (blog). https://www.pewresearch.org/short-reads/2021/04/29/key-facts-about-asian-origin-groups-in-the-u-s/ (accessed July 1, 2024).

Catron, Peter. 2019. "The Citizenship Advantage: Immigrant Socioeconomic Attainment in the Age of Mass Migration." *American Journal of Sociology* 124 (4): 999–1042.

Chan, Sucheng. 1984. "Chinese Livelihood in Rural California: The Impact of Economic Change, 1860–1880." *Pacific Historical Review* 53 (3): 273–307.

Chen, Joyce J. 2015. "The Impact of Skill-Based Immigration Restrictions: The Chinese Exclusion Act of 1882." *Journal of Human Capital* 9 (3): 298–328.

Chen, Shuo, and Bin Xie. 2024. "Institutional Discrimination and Assimilation: Evidence from the Chinese Exclusion Act of 1882." *Explorations in Economic History* 94:101615.

Chew, Kenneth, Mark Leach, and John M. Liu. 2009. "The Revolving Door to Gold Mountain: How Chinese Immigrants Got around US Exclusion and Replenished the Chinese American Labor Pool, 1900–1910." *International Migration Review* 43 (2): 410–30.

Chew, Kenneth S. Y., and John M. Liu. 2004. "Hidden in Plain Sight: Global Labor Force Exchange in the Chinese American Population, 1880–1940." *Population and Development Review* 30 (1): 57–78.

Chin, Gabriel J., and Sam Chew Chin. 2022. "The War against Asian Sailors and Fishers." *UCLA Law Review* 69: 572–622.

Chin, Gabriel J., and John Ormonde. 2018. "The War against Chinese Restaurants." *Duke Law Journal* 67: 681–741.

Chinn, Thomas W., Him Mark Lai, and Philip P. Choy. 1969. *A History of the Chinese in California: A Syllabus*. Chinese Historical Society of America.

Choy, Catherine Ceniza. 2003. *Empire of Care: Nursing and Migration in Filipino American History*. Duke University Press.

Daniels, Roger. 1990. *Asian America: Chinese and Japanese in the United States since 1850*. University of Washington Press.

Eriksson, Katherine, and Zachary Ward. 2019. "The Residential Segregation of Immigrants in the United States from 1850 to 1940." *Journal of Economic History* 79 (4): 989–1026.

Fernandez, Marilyn, and Kwang Chung Kim. 1998. "Self-Employment Rates of Asian Immigrant Groups: An Analysis of Intragroup and Intergroup Differences." *International Migration Review* 32 (3): 654–81.

Flippen, Chenoa A., and Dylan Farrell-Bryan. 2021. "New Destinations and the Changing Geography of Immigrant Incorporation." *Annual Review of Sociology* 47: 479–500.

Frey, William H. 2022. "Today's Suburbs Are Symbolic of America's Rising Diversity: A 2020 Census Portrait." Brookings Institution, June 15. https://www.brookings.edu/articles/todays-suburbs-are-symbolic-of-americas-rising-diversity-a-2020-census-portrait/ (accessed September 20, 2025).

Gonzales, Juan L., Jr. 1986. "Asian Indian Immigration Patterns: The Origins of the Sikh Community in California." *International Migration Review* 20 (1): 40–54.

Grossman, Daniel, Umair Khalil, and Laura Panza. 2025. "The Intergenerational Health Effects of Forced Displacement: Japanese American Incarceration during WWII." *Journal of Public Economics* 242: 105308.

Hanna, Mary, and Jeanne Batalova. 2021. "Immigrants from Asia in the United States." Migration Policy Institute, March 10. https://www.migrationpolicy.org/article/immigrants-asia-united-states-2020.

Hayashi, Kristin. 2021. "The Return of Japanese Americans to the West Coast in 1945." National World War II Museum. https://www.nationalww2museum.org/war/articles/return-japanese-americans-west-coast-1945 (accessed on August 15, 2025).

Higham, John. 1955. *Strangers in the Land: Patterns of American Nativism, 1860–1925*. Rutgers University Press.

Horowitz, Juliana, Ruth Igielnik, and Rakesh Kochhar. 2020. "Most Americans Say There Is Too Much Economic Inequality in the US, but Fewer than Half Call It a Top Priority." Pew Research Center, January 9. https://www.pewresearch.org/social-trends/2020/01/09/most-americans-say-there-is-too-much-economic-inequality-in-the-u-s-but-fewer-than-half-call-it-a-top-priority/ (accessed September 20, 2025).

Hwang, Jackelyn, and Tyler W. McDaniel. 2022. "Racialized Reshuffling: Urban Change and the Persistence of Segregation in the Twenty-First Century." *Annual Review of Sociology* 48: 397–419.

INS (Immigration and Naturalization Service). 2022. *Estimates of the Unauthorized Immigrant Population Residing in the United States*. Office of Homeland Security Statistics, US Department of Homeland Security. https://ohss.dhs.gov/topics/immigration/illegal/population-estimates (accessed August 1, 2024).

Jacobson, Matthew Frye. 2000. *Barbarian Virtues: The United States Encounters Foreign Peoples at Home and Abroad, 1876–1917*. Hill and Wang, Farrar, Straus and Giroux.

Jensen, Eric. 2015. "China Replaces Mexico as the Top Sending Country for Immigrants in the United States." US Census Bureau (blog), May 1. https://www.census.gov/newsroom/blogs/research-matters/2015/05/china-replaces-mexico-as-the-top-sending-country-for-immigrants-to-the-united-states.html.

Kim, Ann H., and Michael J. White. 2010. "Panethnicity, Ethnic Diversity, and Residential Segregation." *American Journal of Sociology* 115 (5): 1558–96.

Kim, ChangHwan. 2025. "Culture and Immigrant Selectivity in Shaping Asian American Education: Evidence from Historical Census Data." *American Sociological Review* 90 (4): 626–57.

Kim, ChangHwan and Andrew Taeho Kim. 2024. "Persistent Educational Advantages of Asian Immigrants'Children, 1940 to 2015–2019." *International Migration Review* 58 (3): 1–33.

Kochhar, Rakesh, and Anthony Cilluffo. 2018. "Income Inequality in the US Is Rising Most Rapidly among Asians." Pew Research Center, July 12. https://www.pewresearch.org/social-trends/2018/07/12/income-inequality-in-the-u-s-is-rising-most-rapidly-among-asians/ (accessed September 30, 2025.

Kye, Samuel. 2023. "The Rise of Asian Ethnoburbs: A Case of Self-Segregation?" *Sociology of Race and Ethnicity* 9 (3): 377–94.

Lee, Erika. 2003. *At America's Gates: Chinese Immigration during the Exclusion Era, 1882–1943*. University of North Carolina Press.

Lee, Erika. 2015. *The Making of Asian America: A History*. Simon and Schuster.

Lee, Hannah. 2021. "Understanding Neighborhood Patterns across Asian Ethnic Groups and Immigrant Destinations." *Spatial Demography* 9 (3): 413–42.

Lee, Heather R. 2014. "Entrepreneurs in the Age of Chinese Exclusion: Transnational Capital, Migrant Labor, and Chinese Restaurants in New York City, 1850–1943." PhD diss., Brown University.

Lee, Jennifer, Kimberly Goyette, Xi Song, and Yu Xie. 2024. "Presumed Competent: The Strategic Adaptation of Asian Americans in Education and the Labor Market." *Annual Review of Sociology* 50: 455–74.

Lee, Jennifer, and Tiffany J. Huang. 2021. "Reckoning with Asian America." *Science* 372 (6537): 8.

Lee, Jennifer, and Karthick Ramakrishnan. 2021. "From Narrative Scarcity to Research Plenitude for Asian Americans." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 7 (2): 1–20.

Lee, Jennifer, and Min Zhou. 2015. *The Asian American Achievement Paradox*. Russell Sage Foundation.

Lew-Williams, Beth. 2014. "Before Restriction Became Exclusion: America's Experiment in Diplomatic Immigration Control." *Pacific Historical Review* 83 (1): 24–56.

Lew-Williams, Beth. 2018. *The Chinese Must Go: Violence, Exclusion, and the Making of the Alien in America*. Harvard University Press.

Lew-Williams, Beth. 2025. *John Doe Chinaman: A Forgotten History of Chinese Life under American Racial Law*. Harvard University Press.

Li, Wei. 1998. "Anatomy of a New Ethnic Settlement: The Chinese Ethnoburb in Los Angeles." *Urban Studies* 35 (3): 479–501.

Lin, Peter Z., and Giovanni Peri. 2025. "How the 1942 Japanese Exclusion Impacted U.S. Agriculture." National Bureau of Economic Research Working Paper Series No. 33971.

Liu, John M., Paul M. Ong, and Carolyn Rosenstein. 1991. "Dual Chain Migration: Post-1965 Filipino Immigration to the United States." *International Migration Review* 25 (3): 487–513.

Logan, John R., and Weiwei Zhang. 2013. *Separate but Equal: Asian Nationalities in the US*. US2010 Project.

Ngai, Mae M. 2004. *Impossible Subjects: Illegal Aliens and the Making of Modern America*. Princeton University Press.

Pascoe, Peggy. 2009. *What Comes Naturally: Miscegenation Law and the Making of Race in America*. Oxford University Press.

Portes, Alejandro, and Rubén G. Rumbaut. 2006. *Immigrant America: A Portrait*. University of California Press.

Postel, Hannah M. 2023. "Record Linkage for Character-Based Surnames: Evidence from Chinese Exclusion." *Explorations in Economic History* 87: 101493.

Postel, Hannah M. 2024. "Summaries of Doctoral Dissertations: 'Records of Exclusion: Chinese Immigration in Historical Perspective.'" *Journal of Economic History* 84 (2): 598–603.

Postel, Hannah M. 2025. "Beyond Urban Chinatowns: Spatial Patterns of Chinese Residence in Pre-Exclusion California." Unpublished.

Postel, Hannah M., and Beth Lew-Williams. 2025. "Beyond Exclusion: The Anti-Chinese Legal Regime in the American West." Unpublished.

Sakamoto, Arthur, Kimberly A. Goyette, and ChangHwan Kim. 2009. "Socioeconomic Attainments of Asian Americans." *Annual Review of Sociology* 35: 255–76.

Saxenian, AnnaLee. 1999. *Silicon Valley's New Immigrant Entrepreneurs*. Public Policy Institute of California.

Singer, Audrey. 2004. "The Rise of New Immigrant Gateways." Brookings Institution, February 1. https://www.brookings.edu/articles/the-rise-of-new-immigrant-gateways/.

Starr, Paul D. 1981. "Troubled Waters: Vietnamese Fisherfolk on America's Gulf Coast." *International Migration Review* 15 (1–2): 226–38.

Takaki, Ronald. 1998. *Strangers from a Different Shore: A History of Asian Americans, Updated and Revised Edition*. Little, Brown and Company.

Taylor, P., and N. G. Ruiz. 2017. "Immigration pathways of Asian immigrants to the U.S." In *The Rise of Asian Americans*, Ch. 2, 29–52. Pew Research Center. https://www.pewresearch.org/social-trends/2017/09/08/immigration-pathways-of-asian-immigrants-to-the-u-s (accessed on August 15, 2025).

Tian, Ziyao, Xi Song, and Yu Xie. 2023. "Intergenerational Social Mobility of Asian Americans under the Shadow of Asian Exclusion (1882–1943)." Unpublished.

Tran, Van C. 2024. "Asian American Diversity and Growth." *Annual Review of Sociology* 50: 581–601.

Tran, Van C., Jennifer Lee, Oshin Khachikian, and Jess Lee. 2018. "Hyper-selectivity, Racial Mobility, and the Remaking of Race." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 4 (5): 188–209.

US Department of State. 2025. *Nonimmigrant Visa Statistics.* Travel.State.Gov. https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/nonimmigrant-visa-statistics.html (accessed September 25, 2025).

Waldinger, Roger. 1996. "From Ellis Island to LAX: Immigrant Prospects in the American City." *International Migration Review* 30 (4): 1078–86.

Warren, Robert. 2021. "In 2019, the US Undocumented Population Continued a Decade-Long Decline and the Foreign-Born Population Neared Zero Growth." *Journal on Migration and Human Security* 9 (1): 31–43.

White, Michael. J., Ann E. Biddlecom, and Shenyang Guo. 1993. "Immigration, Naturalization, and Residential Assimilation among Asian Americans in 1980." *Social Forces* 72 (1): 93–117.

Zeng, Zhen, and Yu Xie. 2004. "Asian-Americans' Earnings Disadvantage Reexamined: The Role of Place of Education." *American Journal of Sociology* 109 (5): 1075–108.

Zhou, Min. 1992. *Chinatown: The Socioeconomic Potential of an Urban Enclave.* Temple University Press.

Zhou, Min, and Carl L. Bankston. 1998. *Growing Up American: How Vietnamese Children Adapt to Life in the United States.* Russell Sage Foundation.

Journal of Economic Perspectives—Volume 40, Number 1—Winter 2026—Pages 215–240

From Asia, with Skills

Gaurav Khanna

Migration from countries in Asia in the last three decades or so—and especially China and India, as well as South Korea, the Philippines, and Japan—has occurred with unprecedented scale and impact. This wave of migration to the United States has been predominantly high-skill. In 2019, among working adults in the US labor force, 78 percent of Indian-born and 63 percent of Chinese-born workers held at least a four-year college degree, compared to 39 percent of US-born workers (ACS 2019). Between 1990 and 2019, the share of college-educated workers in the United States from these five Asian countries doubled to 7.3 percent. Over this period, migration from these countries contributed to over 38 percent of the growth in US employment of software developers, 25 percent of the increase in scientists and engineers, and 21 percent of the growth in physicians (ACS 2019).

In this article, I examine the rise in high-skill migration from Asia. I begin with an investigation of broader trends, and how we got here. It focuses on the underlying reasons for migration from five Asian countries: India, China, South Korea, Japan, and the Philippines. I discuss the existing literature and use census data, visa records, patent counts, and surveys to explore both US demand-side factors and Asian supply-side dynamics that contribute to these trends. Concurrently, I examine the broader economic impacts of this migration on the US economy—particularly on key sectors such as information technology, which play a disproportionate role in driving innovation and growth both directly and mediated through downstream industries like manufacturing that rely on these technologies as inputs.

■ *Gaurav Khanna is Associate Professor of Economics, School of Global Policy and Strategy, University of California—San Diego, San Diego, California.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251454.

Together, these findings underscore the significance of visa policies, the student-to-skilled worker pipeline, and the contributions of Asian-born professionals to innovation and productivity. I conclude by looking ahead at several factors that will determine whether the United States can continue to attract and utilize talent from Asia in the face of increasing competition from the rest of the world, which in turn will shape the countries in which the next technological innovations, in sectors from artificial intelligence to pharmaceuticals, are most likely to occur.

Economic and Policy Drivers of Skilled Migration

I begin by documenting the major trends in high-skill immigration from Asia and the push and pull factors that facilitated them. I start with US economic and policy changes, and then explain why Asian migrants were particularly well-positioned to address these changes.

A Shifting Need for Skills in the US Economy

Domestic policy, market forces, and demographic pressures have changed the need for skills across the US economy. The information technology sector is a primary example, in which the US economy experienced a surge in demand for skilled labor in the 1990s. Nationwide commercial traffic on the internet surged in the mid-1990s, especially following the decommissioning of the National Science Foundation Network in April 1995 (Leiner et al. 1997). The entry and growth of tech firms, like Yahoo, Amazon, and eBay, contributed to the information technology boom of the 1990s. This boom dramatically increased the demand for software developers and computer scientists.

While overall US employment grew by about 28 percent between 1990 and 2019, computer science workers grew fourfold from about 1 million workers in 1990 to 4.34 million in 2019 (ACS 2019). These increases in employment were concurrent with substantive increases in wages, which suggests increased demand for computer science and information technology workers. This was likely driven by the rapid rise in internet usage for commercial purposes and immigrant-led innovation or labor market complementarities that amplified demand in the sector.

Technological innovation and entrepreneurship in the information technology sector had broader impacts on tech-adjacent industries as well. The National Science Foundation (2019) estimates that in the 2000s, computer technology patents were the largest contributor to US patenting activity (followed by patents in digital communication), and the fastest-growing. Much of the value-added growth in the early 2000s was driven by industries using information technology (Jorgenson, Ho, and Samuels 2016). US firms increasingly sought scientists and engineers—whether trained domestically or internationally—to drive innovation in a wide range of industries.

At the same time, the US higher education sector underwent some dramatic changes over the last few decades. State appropriations for public universities had

been declining since the 1980s, with particularly dramatic decreases during the recessions in the early 1990s, the dot-com bust in 2001, and the Great Recession in 2008. In this setting, the availability of international students changed the revenue model of US higher education, as the US higher education sector shifted to "exporting services" to an international consumer base. One especially visible manifestation was the introduction of many new master's programs in fields related to science, technology, engineering, and mathematics (STEM) (Bound et al. 2020, 2021). The rise in students from China was particularly pronounced at public research universities between 2005 and 2016, after which there was a rapid growth in students from India. As US universities catered to student demand from abroad, there was a concurrent increase in demand for foreign-born faculty. In turn, many of these international students would transition into the US labor force, particularly in high-tech sectors.

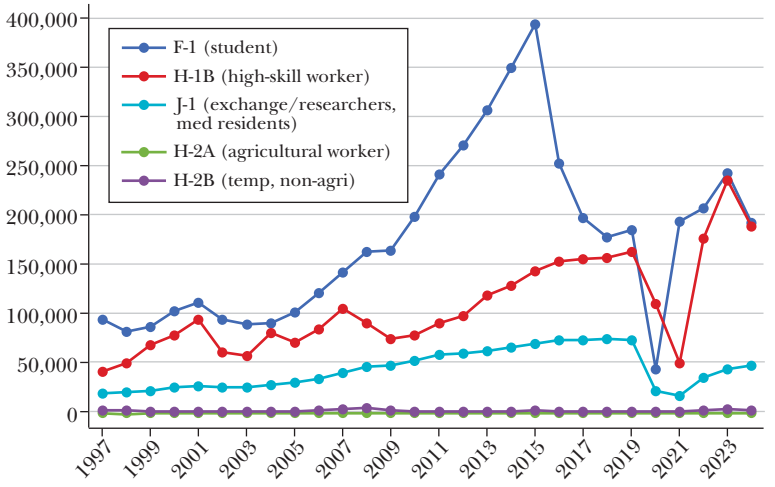
Concurrently, demographic aging in the United States has sharply increased healthcare utilization over the last few decades, placing sustained pressure on the supply of physicians, nurses, and long-term care providers (De Nardi et al. 2015; Jung, Tran, and Chambers 2017; Auerbach, Buerhaus, and Staiger 2020). Yet, the United States has also simultaneously imposed policy restrictions on the production of US-educated medical professionals, including a moratorium on new medical schools, fixing class size, capping federal funding for residencies, and freezing residency positions (Orr 2020). By the early 2000s, US healthcare establishments perceived a shortfall of physicians and nurses (AAMC 2021) and increased their demand for medical residents, physicians, and nurses from abroad. Changes in US immigration policy that facilitated the migration of physicians, particularly for residency programs, then established a pathway to the US labor force.

US Immigration Policy

The history of migration from Asia was marked by exclusionary laws and systematic barriers (as Postel discusses in this issue). This situation began to change with the 1965 Immigration and Nationality Act and the removal of country-specific barriers. The post-1965 system emphasized family sponsorship, but generations of historical Asian exclusion meant few family sponsors. Thus, between 1965 and 1990, many immigrants entering the US labor market were originally granted entry via permanent residency or "green card" provisions. Employment-based green cards require US firms to petition to hire an immigrant worker and certify that they cannot hire a US citizen for the particular position. Most employment-based green cards, which are still primarily governed by the 1965 Act, are currently granted through an adjustment of status for individuals already in the United States.

But starting in the early 1990s, US immigration policy shifted to attracting high-skill immigrants under "temporary" status. As a result, most of the new high-skill migrants from Asia initially entered under temporary work status with the H-1B visa, under student status with the F visa, or as exchange visitors (including scholars and physicians) with a J-1 visa. As a result, entering on a temporary work or student visa is the first step to entry for most high-skill migrants from Asia. In 2010, temporary work visas accounted for 39 percent of the first visa status of information

Figure 1
Visas Granted to Citizens of India, China, South Korea, Japan, and the Philippines



Source: Author's calculations using data from US Department of State (2025).
Note: F-1 visas apply to students, H-1B visas for high-skill workers, J-1 visas for exchange researchers/medical residents, and H-2A/B visas for agricultural and non-agricultural work, respectively. Data spans Fiscal Years 1997 to 2024.

technology workers aged 25–34, and student visas accounted for about 35 percent (Bound et al. 2015b). In contrast, older cohorts were more likely to have entered through permanent residency.

Figure 1 shows visas granted to individuals from the five Asian countries. Several factors stand out. First, visas are predominantly issued to college students and graduates. Visas for students (F-1), high-skill workers (H-1B), and researchers/medical residents (J-1) vastly outnumber those for low-skill agricultural and nonagricultural work (H-2A and H-2B). Second, starting in 2005, there was a sharp increase in F-1 student visas, peaking in 2015. As discussed later, the rapid growth in the 2000s and the slump after 2016 were both primarily driven by Chinese nationals.[1] The sharp declines following the Covid-19 pandemic were followed by a resurgence in student flows, this time fueled by Indian nationals. Finally, H-1B visas have experienced steady growth since the 1990s; however, due to Congressional caps on the visas, this growth has been muted compared to student visas as a pathway. As discussed below, H-1B visas were predominantly granted to information technology workers from India. Taken together, visa policies, along with the demand from US universities for international students and from the information technology sector for tech workers, played a significant role in driving the skill bias in recent Asian migration.

[1] The 2017 decline was partly driven by a change in the requirements for English Language Programs (ICE 2016).

Certain key features of US immigration policy pointedly aim to attract specific migrants in distinct sectors of the economy. For example, student visas are uncapped and do not require a costly petition from a US-based employer, unlike most other visa categories, and so student visas have the largest room to grow. In addition, graduating from a US university allows workers to join the US labor force temporarily through the OPT (Optional Practical Training) program. Each degree (bachelor's, master's, PhD) allows the student to work in the United States on an OPT for one year, adding an additional two years if the degree is in a field related to science, technology, engineering, or mathematics. The OPT program further encouraged Asian students to choose the science-related fields once in the United States (Anelli, Shih, and Williams 2023). In addition, US immigration policy grants an extra 20,000 H-1B visas for those with US graduate degrees, again incentivizing students to obtain a graduate degree from a US institution and then stay on in the US workforce. Together, these factors enable many Asian migrants to view US higher education as a stepping-stone into the US labor force.

US immigration policy is also targeted to certain areas and industries. For example, the H-1B visa is capped by a numerical amount determined by the US Congress, but includes an uncapped exception for foreign-born scientists, faculty, and researchers at US nonprofit universities and research centers. Furthermore, the H-1B visa is demand-driven, in that firms and organizations must petition for workers (unlike many other countries, where skills determine visa success, often without requiring concrete job offers). The H-1B visa is restricted to "specialty occupations," and so employment must be in certain high-skill jobs that require "theoretical and practical application of a body of highly specialized knowledge."

This needs-based focus also manifests in other aspects of high-skill migration policy. Visas are more available for entrepreneurs, managers within multinational firms, and for certain specialized service occupations, such as medical services. For instance, to meet the shortfall of US physicians and nurses, immigration policy was tailored to attract medical professionals to stay and work in high-need areas. US immigration policy for nurses operates through "Schedule A" visa designations, which identify occupations facing chronic domestic shortages and therefore exempt employers from the labor certification process normally required for employment-based green cards. Under this provision, hospitals and healthcare institutions can directly sponsor foreign nurses—many from the Philippines—without demonstrating the unavailability of US workers.

Immigration policy also mediates the pathway from medical residency to clinical practice. Traditionally, medical residents on J-1 visas must return to their home country for at least two years before they can apply for a temporary work visa in the United States. However, the Conrad 30 Visa Program was established in 1994 to waive this home residency requirement as long as physicians transition to a longer-term work visa at the end of residency in a "medically under-served area" or "health professional shortage area." To receive a waiver, a foreign physician must be recommended by an interested government agency. Since 2001, more than 18,000 foreign physicians have participated in the program, concentrated in rural and underserved

areas. As Braga, Khanna, and Turner (2024) show, the program responds directly to demand pressures: states facing greater physician shortfalls make fuller use of their waiver allocations, thereby channeling foreign-trained doctors toward underserved regions. Once again, visa waivers are concentrated among those from Asia, with the majority of participants coming from the Philippines, India, and Pakistan (Ranasinghe 2015).

Asia's Advantage in Meeting Rising Demand

While US immigration policy facilitated the increase in demand for skilled labor, migrants born in Asian countries had a distinct advantage in meeting this rising demand. During the 1990s, many Asian countries experienced a dramatic increase in the number of young people completing high school and undergraduate college education. For instance, China's 211 Project dramatically increased the size of the higher education sector. Also in China, the gross enrollment ratio in tertiary education expanded from 3 percent to 75 percent between 1990 and 2023, while in India, this ratio grew from 6 percent to 33 percent (World Bank 2025). In Korea, the gross enrollment ratio expanded from 33 percent in 1990 to 100 percent by 2021. Assuming that talent and ability distributions are similar in the United States and abroad and given the large number of college-ready and college-educated youth in South and East Asia, the quantity of potential talent available to US firms and institutions was quite large. For perspective, India has more young adults between 20 and 34 years of age than the entire US population.

In many Asian countries, such as India, China, and the Philippines, the returns to skill remain relatively limited, making US employment especially attractive. Clemens (2013) uses personnel records from a large Indian multinational information technology firm to compare the earnings of H-1B lottery winners and losers. Exploiting the random allocation of visas, the paper estimates a causal return to moving a software developer from India to the United States. They find a six-fold increase in earnings (after converting earnings at the market exchange rate), highlighting the substantial wage premium associated with US employment. These findings underscore the high returns to human capital in the United States—potentially driven by technological complementarities and location-specific spillovers. The ability of US firms to access a vast pool of highly skilled workers from Asia at comparatively lower wages further enhances the appeal of these workers to American employers.

In a simple Roy (1951)–style framework, the above factors would enable Asian countries to dominate the flow of high-skill migrants to the United States; that is, a combination of limited opportunities for students and workers at home, combined with a large number of talented youth in the right tails of the aptitude distribution. However, other significant events reinforced these trends.

The rapid income growth in Asian countries in the 1990s increased the affordability of overcoming the cost of migration. For example, China's per capita GDP (in purchasing power parity-adjusted constant 2021 dollars) has increased more than 13-fold since 1990. The manufacturing boom in China after it joined the World Trade Organization in 2001 led to substantial income gains and, consequently,

increased student migration to the United States. Khanna et al. (2023) use detailed data on the universe of foreign students to show that households in Chinese cities that grew as a result of favorable tariffs sent more youth to the United States, as they were now more able to afford a US higher education degree.

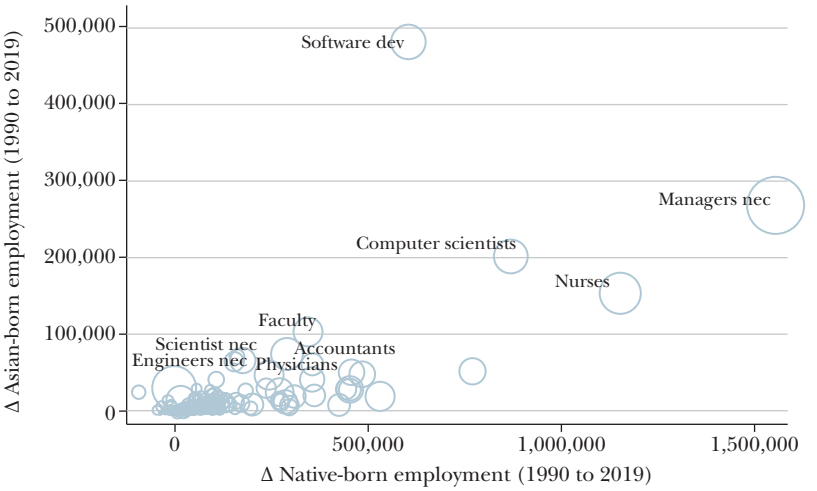
In addition, economies across Asia were also making domestic investments in high-quality science and engineering institutions, which enabled a growing number of students to meet the skill requirements of US-based jobs. The first Indian Institutes of Technology were established in the 1950s and are among the highest-return universities in the world, based on the amount that they increase wages in a common labor market (Martellini, Schoellman, and Sockin 2024, Table B4). Also, in India and across South Asia, the urban population was widely comfortable and trained in English. China has aimed to replicate the success of high-quality institutions abroad through substantial increases in funding and infrastructure through its 211 Project and Project 985. Korean and Japanese technological universities were established to drive growing local high-tech industries. Many of these skills may also be complementary to the education offered in US institutions (or workers trained in Canada, the United Kingdom, or Australia), and US firms may seek such talent to work in teams with their own locally educated workforce (Peri, Shih, and Sparber 2015).

Similarly, the Philippines has long held an advantage in supplying nurses abroad. Specifically, its nursing education system was modeled on the US system during the twentieth century, with English-language instruction, US-aligned curricula, and licensure standards designed to meet international requirements. This institutional alignment made Filipino graduates particularly attractive to US employers under the Schedule A visa category. Moreover, Abarcar and Theoharides (2024) document that the Philippines has an extensive network of nursing schools, many established to serve the export market, which expanded rapidly in response to overseas demand. As a result, US policy shocks—such as visa expansions or increased hospital demand—translate quickly into higher domestic nursing enrollment and training capacity in the Philippines.

Over time, the professional and alumni networks of these institutions gave emigrants from Asian countries a further advantage. For instance, India had sent a number of top engineers during the earlier hardware boom of the 1980s, and the diaspora helped establish strong connections and a reputation for well-trained workers (Saxenian 1999). As Bhatnagar (2006) notes, Indian professionals in Silicon Valley built networks and established strong reputations, which they then leveraged to help attract the next wave of Indian migrants to the expanding information technology sector within US firms. The reputation was further bolstered by the offshoring of low-level tasks during the Y2K crisis—the months leading up to January 2000, when there was concern that the change in annual dates might crash some computer systems—when US firms outsourced work to India, finding that the twelve-hour time lag allowed US companies to send work at the end of the US workday and pick it up as the workday ended in India (Arora and Athreye 2002).

To some extent, this rapid late-twentieth-century rise in Asian migration to the United States also reflects the possible release of long-suppressed demand created by

Figure 2
Change in Number of College-Graduate Workers between 1990 and 2019, by Native-Born and Asian-Born Employment



Source: 1990 US Census and 2019 ACS.
Note: IPUMS occ90 occupation codes, "nec" is "not elsewhere classified." "Faculty" includes HS/college subject instructors.

decades of exclusionary policy. For much of US history, immigration from Asia was effectively prohibited: naturalization was barred from 1871 to 1952, and entry was restricted for those "ineligible for citizenship" from 1917 onward (as discussed by in this issue by Postel). When these legal and network constraints were finally relaxed, migration rose rapidly from a very low base. The long history of exclusion thus also shaped both the timing and the high-skill composition of Asian immigration.

Major Occupations of Asian Migrants: The Role of Demand and Supply

The fastest-growing occupation among Asian-born immigrants from 1990 to 2019 was software development, followed by managers, computer scientists, nurses, faculty, scientists, accountants, engineers, and physicians, as shown in Figure 2, using Census data on college graduates in the US labor force. A natural question is: Does this high growth in specific occupations primarily reflect the US demand or supply-side push factors from Asia? If supply-side forces from Asia dominate, then we might expect that Asian migrants "crowd out" US-born workers in these specific high-skill occupations; on the other hand, if occupation-specific US demand is dominant and the supply of local talent is constrained, then crowding out of US-born counterparts is less likely to occur.

Of course, both demand and supply forces are at play. As noted earlier, various constraints have limited the supply of US-born workers in health care industries, and while there were no similar restrictions on scientists, engineers, and computer

scientists, acquiring such skills can be challenging and costly. Moreover, US immigration rules have limited the supply of skilled labor. But overall, the fact that Figure 2 shows a roughly upward-sloping relationship between the growth in native-born and Asian-born employment suggests that US demand-side factors might be important; that is, occupations with growing US native-born employment are also areas with growing Asian immigrant employment. For instance, innovation in the tech sector increased the demand for all software developers, whether born in the United States or Asia, resulting in growth for both US-born and Asian-born software developers. This upward-sloping relationship may also reflect immigrant-led innovation (Kerr and Lincoln 2010) and complementarities in production that increase the demand for native-born employment (Peri, Shih, and Sparber 2015). Had the relationship in Figure 2 been downward sloping, it would be consistent with Asian-born programmers displacing US-born ones rather than meeting a growing demand. Of course, this is not to say the growth in US-born programmers (or their wages) might not have been higher in the absence of supply from Asia.

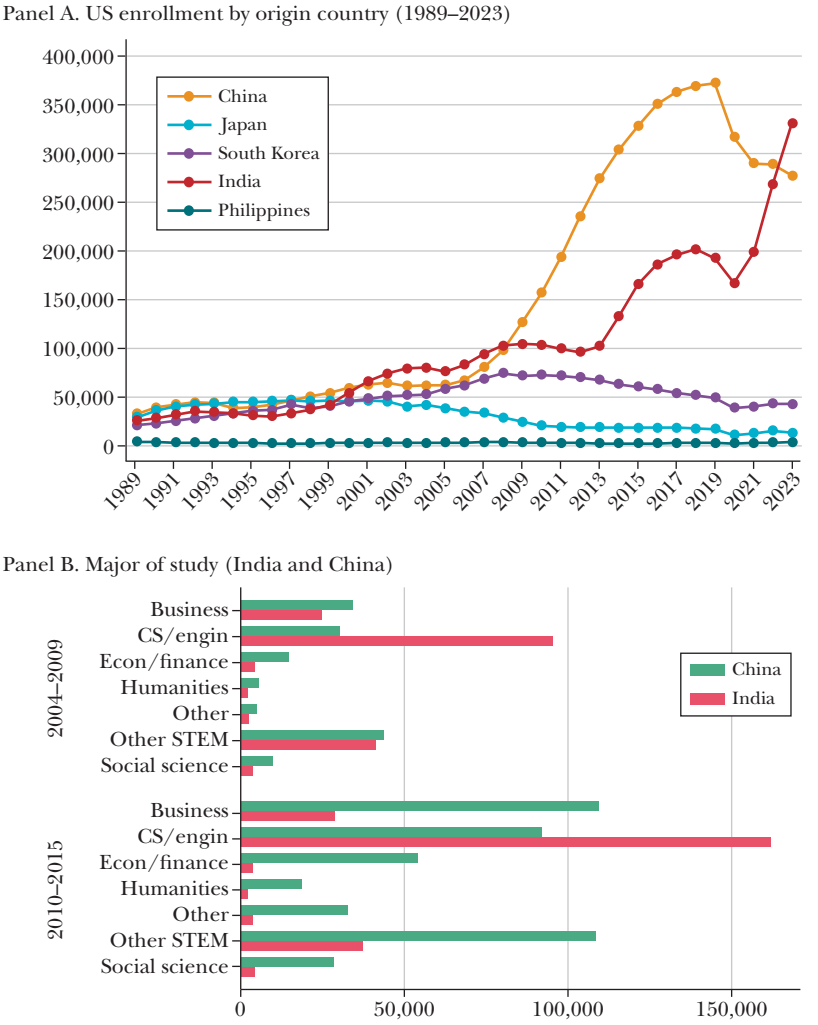
In subsequent sections, I build from the patterns shown in Figures 1 and 2 to study the following groups in detail: students and faculty in higher education, software developers and computer scientists (the information technology boom), managers and scientists (entrepreneurship and innovation), and nurses and physicians (medical professions). For each group, I dig deeper into the demand-side factors from the US economy, why migrants from Asian countries were well-suited to meet this demand, and the implications and downstream impacts for the economies sending and receiving skilled migrants.

The Higher Education Sector: Students and Faculty

The number of international student visas is not capped—unlike work visas—and so migrants can use them as a stepping stone into the high-tech US labor market. In 2010, 27 percent of all foreign-born workers and 35 percent of foreign-born information technology workers in the United States initially arrived on a student visa (Bound et al. 2015b). Asian-born students are heavily concentrated in science and engineering fields, such as computer science, engineering, and mathematics.

Figure 3, panel A, plots the trends in US international student enrollment by foreign country of origin. As this figure shows overall enrollment (a stock measure that includes seniors in college), these trends lag behind new visas issued to freshmen students (a measure of new flows). A few changes in patterns over time stand out. First, enrollment from China started increasing in the mid-2000s and grew rapidly until about 2016. The slowdown was followed by a sharp decline in Chinese enrollment during the Covid-19 pandemic, which never really recovered. The initial growth and slowdown in enrollment lag the visas granted (Figure 1), which is a measure of more immediate changes to demand and new flows. Second, enrollment from India shows a steady rise, which increases more rapidly after 2014 and then

Figure 3
International Students at US Colleges and Universities



Source: Panel A: Author's calculations using Open Doors Data (Institute of International Education). Panel B: Author's calculations using visa data obtained through a FOIA request (USCIS 2017).
Note: Panel A shows international students in the United States, by country of origin (1989–2023). Years represent the start of the academic year. In panel B, numbers represent the sum of all new visas granted during a period. "CS/Engin" majors include computer science and engineering, while "Other STEM" excludes that category. STEM stands for "science, technology, engineering, and mathematics."

jumps up again after the end of the Covid-19 pandemic. Third, while not shown in the figure, the large initial increase in enrollment from China was for undergraduate degrees, whereas enrollment from India was initially in master's programs and, only recently, also certain undergraduate programs. Finally, although other Asian

countries were substantial sources of international students in the 1990s, they did not keep pace with the growth in flows from China and India.

This rise in US international students coincided with US colleges needing to enroll more foreign students, as noted earlier. The corresponding growth in enrollment from Asia was concentrated in large, non-capacity-constrained public universities, which expanded their enrollment in engineering and in science, technology, engineering, and mathematics fields. Many of these colleges, such as Purdue, Michigan State, Ohio State, Penn State, and Indiana, were in the US Midwest. Around the Great Recession, between 2007 and 2012, public research universities experienced a 133 percent increase in foreign first-time undergraduate enrollment, while private research universities experienced a 61 percent increase.

Figure 3, panel B, focuses on college majors for the two main sending nations, India and China, and compares the 2004–2009 period to the 2010–2015 period. In the early period of 2004–2009, India focused more on computer science and engineering majors, while Chinese students were relatively more concentrated in business and other STEM majors. During the subsequent growth period, Chinese students in business, computer science/engineering, and STEM majors experienced uniform and substantial growth. For Indian students, there was little growth in business and other science majors, while the computer science and engineering majors grew rapidly. These patterns of major specialization have subsequent downstream impacts on the labor force after these students graduate. In the 2023/2024 academic year, 23 percent of Chinese students and 43 percent of Indian students enrolled in math/computer science majors (Institute of International Education 2024). For Japanese students, the largest major was business (18.3 percent), and for Koreans, it was engineering (17 percent).

What Drives Growth in International Student Enrollment?

Why has the surge in international students in US higher education since 2005 come primarily from Asia, and in particular from China and India? In the case of China, some key factors have already been mentioned. US higher education was more broadly affordable in China after it joined the World Trade Organization in 2001, which led to an export-driven boom in incomes, especially after China allowed the renminbi yuan to appreciate in 2005. In addition, China's uniquely rapid expansion of its own undergraduate education led to a surge in graduate students going abroad several years later (Jia et al. 2025). China's 211 Project, launched in 1999, upgraded universities and colleges and expanded undergraduate enrollment. Over the next two decades, the number of universities in the country increased from 1000 to 2900 and the number of students enrolled grew from 1.1 million to more than 7.9 million, and this drove a surge of graduate students to US programs in science, technology, engineering, and mathematics (Jia et al. 2025).

Enrollment of Chinese students in foreign higher education grew in several countries. Between 1998 and 2019, Chinese enrollment in Australia increased by a factor of 37, in the United Kingdom by a factor of 41, and in Canada by a factor of 26. But despite the massive growth to other destinations, even in 2019, there were

about as many Chinese students in the United States as in these other three countries combined (UNESCO 2025).

The rapid growth of international students from China halted in 2016 as new student visas fell (Figure 1) and enrollment tapered off (Figure 3, panel A). Political tensions also contributed to the decline in Chinese enrollment. This decline was concentrated in public universities in red states, and particularly in sensitive fields flagged by the US government (Chang et al. 2025). The concurrent expansion of local Chinese universities and of universities in Australia, Canada, and the United Kingdom helped divert these potential students away from US institutions.

In the case of India, enrollment of students in US higher education began to rise dramatically in 2014. Local constraints in the Indian higher education sector led to increasing competition for limited seats at high-quality Indian universities. As a result, a burgeoning youth body that finished high school and could afford to go abroad looked to the United States and the United Kingdom. By 2023, the number of Indian students enrolled in US universities surpassed that of students from China. One distinction between international student flows from India and China is that students from India are more likely to transition to the US labor force. As noted earlier, the Optional Practical Training program allows students to work in the United States after graduation for at least one year, and for up to three years for those with a degree in science, technology, engineering, and mathematics. In general, transition rates from student status to OPT status are inversely correlated with origin-country GDP per capita (Bound et al. 2021). Richer countries, such as Korea, have relatively lower transitions to OPTs, given the opportunities available back home. For India, more than 90 percent of Indian master's students in our administrative visa data switched to the OPT in 2015 (Bound et al. 2021), while the transition rates for students from China were lower, at about 70 percent.

International Researchers and Faculty

International scholars at US universities are also heavily represented by China, India, Japan, and South Korea. These four countries were the top four sources of scholars for the entire first decade of the 2000s, and even in 2024 the top three sources are China, India, and South Korea (Institute of International Education 2024).[2] In 2024, China accounts for 21 percent, India 16 percent, and South Korea 6.6 percent of international scholars on temporary visas.

Many scholars first enter as students and transition to the US higher education sector after completing their graduate degrees and, down the line, contribute to high-quality research and publications in the United States. Increases in foreign PhD students lead to higher publication and patent counts in US science and engineering departments (Stuen, Mobarak, and Maskus 2012). Remember, universities

[2] The Institute of International Education data define international scholars as scholars on nonimmigrant visas engaged in academic activities and not enrolled as students at a US college or university. As many scholars may have transitioned to a permanent resident status, the numbers here are likely an undercount.

are exempt from the H-1B cap, and many scholars from other countries first begin as an OPT before transitioning to an H-1B.

Effects on the US Universities and the Workforce

Student migration has had a major effect on the revenues of US universities. Public universities, facing state funding cuts, offset revenue losses by enrolling full-fee-paying international students. This strategy allowed them to maintain quality, focus on research, and keep in-state tuition low by cross-subsidizing local students (Bound et al. 2020).[3] Between 2010 and 2016, only 6.6 percent of Chinese undergraduates at US research universities received funding from the university, with most students paying full tuition.

In particular, many students from Asia have been enrolled in revenue-generating master's programs. These master's programs are concentrated in science, technology, engineering, and mathematics, with students from India particularly concentrated in computer science master's programs. In 2023, among full-time students in master's programs in science, engineering, and health, temporary visa-holders outnumbered US citizens (Smith et al. 2024). Cross-subsidization of US students works both through revenue sharing and a graduate student workforce engaged in teaching and research, which allows the university to enroll more local undergraduates while maintaining research quality (Shih 2017).

This cross-subsidization of the local students was made explicit in a public letter from University of California then-president Janet Napolitano (2016), which states: "California's situation is not unique. Nearly every state in the nation has faced this Hobson's choice, and they have all reached the same decision: open doors to out-of-state students in order to keep the doors open for in state." Indeed, the US comparative advantage in higher education services led to higher education service exports of $56 billion in 2024 (Bureau of Economic Analysis 2025), further supporting the economies surrounding college towns as well (Jia et al. 2025).

In the long run, as Asian-born students join the US labor force their impact is felt more broadly in the US economy. Many international graduate students eventually become faculty, contributing to research production, patenting, and teaching at US universities. Approximately 85 percent of US PhD graduates from India and China remain for at least ten years (Finn and Pennington 2018). Looking at just US PhDs in science and engineering fields that graduated between 2017 and 2019, 83 percent of Chinese, 86 percent of Indian, but only 50 percent of South Korean graduates were in the United States in 2023 (NCSES 2025).

Many Asian-born graduates transition to high-skill sectors in information technology and healthcare (Bound et al. 2015b), which are examined below. The H-1B separately allocates 20,000 visas for those with US graduate degrees. The

[3] Bound et al. (2020) combine individual-level international student data via requests under the Freedom of Information Act with university-level foreign enrollment data from IPEDS. They exploit shocks to state budgets to show that cuts to higher education funding lead universities to offset revenue losses by enrolling more international students, especially from China.

first transition step is through an Optional Practical Training program, and transition rates vary by country due to differing home opportunities. In particular, over 60 percent of undergraduate and 90 percent of master's students from India move to work under OPT, while this proportion is significantly lower for students from other Asian countries.

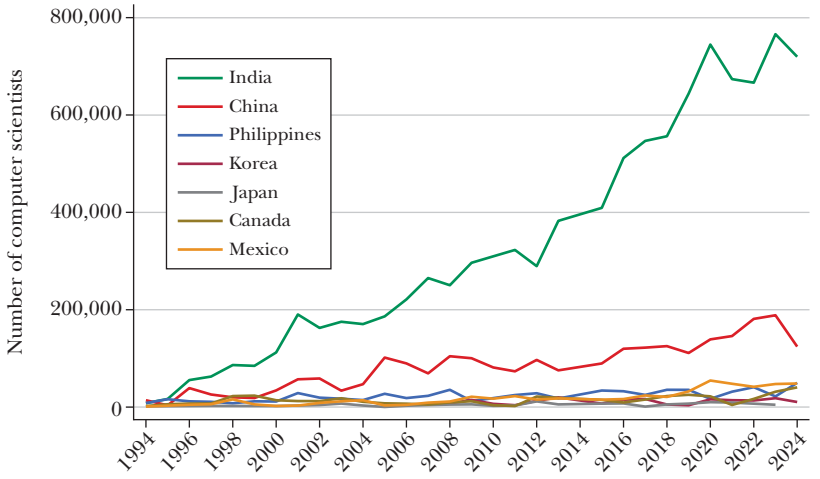
Computer Scientists and the Information Technology Boom

The influence of Asian migrants is especially strong in several industries, including the information technology sector. The rise of Internet commerce made information technology occupations—computer science, software development, and programming—the fastest-growing occupation in the 1990s, more than doubling in size between 1993 and 2010. For example, the share of computer science majors as a fraction of all college graduates in the United States more than doubled between 1990 and 2020, rising from about 2.5 percent to nearly 6 percent. Concurrently, the share of foreign-born workers in computer science occupations also more than doubled, from about 17 percent to 36 percent. As Figure 4 shows, this growth in foreign-born representation came particularly from India: computer science growth from India far outpaced that from China and other countries. By 2010, one-third of US information technology workers were foreign-born, predominantly from India. By 2019, Indian-born individuals were 29 percent of all workers in science, technology, engineering, and mathematics jobs in the United States (AIC 2022).

Why India? Given the sharply rising demand for information technology workers starting in the 1990s, US employers may have found a perfect storm in India's talent pool: as discussed earlier, some main factors include a huge cohort of young, tech-trained, English-speaking workers, with alumni networks and institutional reputations already in place, willing to work for lower (but still life-changing) wages. Other Asian counterparts did not have similar advantages, including English proficiency, diaspora networks, and (compared to Japan and Korea) a wage advantage. This boom coincided with the H-1B visa program, established by the Immigration Act of 1990, which mandated the visa required "theoretical and practical application of a body of highly specialized knowledge in a field of human endeavor." By 2014, 86 percent of all H-1B computer scientists were from India, with 5 percent from China; in 2023, the latest data show that 78 percent of H-1B visas went to Indians (US Department of State 2025).

After this migration, a meaningful fraction of high-skill Indian-born computer scientists reside in the United States: Does this "brain drain" hinder India's economy? The evidence suggests nuance, as the opposite may be more likely. The prospect of migrating to the United States and earning wages that were four to six times higher (Clemens 2013) induced many Indians to accumulate skills and human capital valued abroad—a form of "brain gain," as argued by Khanna and Morales (2025). Indeed, the Indian higher education sector responded to higher demand by expanding

Figure 4
Computer Science Workers by Origin Country (1994–2024)



Source: Author's calculations based on Current Population Survey (CPS) ASEC data (Flood et al. 2024).
Note: Figure shows computer scientists and software developers in the United States, by country of birth (1994–2024). Computer scientists are defined as having occ90 codes 64 and 229. Includes naturalized citizens.

computer science education in top schools. However, because the number of US visas was capped, many who gained skills with the prospect of migrating to the United States remained in India. Furthermore, H-1B visas expire after six years, and the wait for a US green card is both costly and lengthy for Indian nationals. As a result, about 24 percent of Indian visa-holders leave the United States after six years, and many return to India with skills acquired abroad—often after sending remittances back to India during this period. This "brain circulation" complements the initial brain gain and builds a large information technology workforce in India, helping create a highly productive sector with possibilities for innovation spillovers. Indian firms have surpassed the United States in information technology exports, often as more work is offshored to India. In this way, the shift in information technology production from the US economy to the other side of the world in India was partially facilitated by quirks of US immigration policy—specifically, the binding H-1B caps and the costly green card process.

From a US perspective, the influx of Indian tech workers affects both the tech sector and the broader economy. Indian tech workers help expand production from the US information technology sector, but they also facilitate offshoring certain aspects of production to India. For example, this may entail specialization in new product development at US headquarters, but software coding in their India offices.

A large body of research uses matched administrative data on H-1B petitions, worker earnings, and firm employment to estimate the effects of high-skilled immigration on US labor markets (Peri, Shih, and Sparber 2015; Ottaviano and Peri 2012; Doran, Gelber, and Isen 2022; Mahajan et al. 2024; Kerr and Lincoln 2010; Bound et al. 2015a). These studies exploit variation generated by H-1B visa-cap, H-1B lotteries, other policy changes, and differences across firms and regions in H-1B dependence to identify causal impacts on wages, employment, and innovation.

Taking this evidence as a whole, Indian-born computer scientists may lower wages for US-born computer scientists, but more computer scientists also raise wages for complementary jobs where natives are concentrated. As information technology firms hire more Asian-born programmers, the demand for managers and workers in complementary occupations like human resources increases. As a result, wages for these complementary professions may rise and US-born workers may switch into these positions, potentially away from programming tasks. The impact on computer science workers themselves may be ambiguous if immigrant computer science workers are particularly involved in innovation and raise wages broadly across the firm, but still face meaningful competition from more programmers (Bound et al. 2015a). Furthermore, over time, outsourcing companies started winning almost 20 percent of H-1B visas, and they may be paying relatively lower wages than other tech firms. Turner (2022) leverages changes to the H-1B cap to show that a 10 percentage point increase in immigrant share of the workforce in fields of science, technology, engineering, and mathematics lowers wages by 1 percent.

While attention often focuses on whether Asian immigrant workers displace native-born US workers within specific industries, the benefits to consumers and the broader economy have been relatively understudied. Yet some of the largest gains from tech sector immigration may actually be to workers outside the tech sector, as consumers use better software and firms utilizing more information technology inputs see productivity gains (Khanna and Morales 2025). Prices of software, internet publishing, data processing, programming services, and computer systems design fell rapidly in the 1990s and 2000s, while the quantity index of these products saw a massive surge (Bureau of Labor Statistics 2025). Furthermore, many sectors of the US economy (like auto manufacturing and banking) now rely on information technology products as inputs into production or as part of the products they sell (Jorgenson, Ho, and Samuels 2016). These downstream sectors experienced significant productivity gains, far outpacing those of other industries outside the information technology sector.

Entrepreneurship and Innovation

Ultimately, the positive overall economic impact of tech workers in both the United States and India is driven by their high propensity to innovate, create new products, and establish new enterprises, especially in industries that involve science and engineering. Many information technology workers who may first enter the

United States on a student or H-1B visa make a transition from technical roles to becoming entrepreneurs and innovators (Hunt 2011). However, the H-1B visa is not focused on entrepreneurs. In contrast, Canada has a "Startup Visa" that attracts talent, particularly Asian migrants who were previously based in the United States (Lee and Glennon 2023). Australia and other countries have also introduced founder-friendly visas to attract entrepreneurial talent.

Many studies highlight different measures of the roles of Asian immigrants in innovation, including participation in the National Academy of Science, Nobel Prize recipients, and high-quality publications (Chellaraj, Maskus, and Mattoo 2008; Black and Stephan 2010; Gaulé and Piacentini 2013). Other studies draw on patent-level data, inventor name matching, and firm-level administrative records, exploiting temporal and geographic variation in H-1B admissions and foreign-student inflows to isolate how skilled immigration drives innovation, patenting, and firm growth. For example, innovation in industry is often measured by patenting behavior. The US Patent and Trademark Office maintains records of names, and US patents have shown sharp increases in the use of ethnically Chinese and Indian names in US patents (Kerr 2008). Skilled immigrants (in this period, mostly from Asia) also appear to have an advantage over natives in patenting and publishing (Hunt 2011) and boost innovation more broadly (Hunt and Gauthier-Loiselle 2010; Kerr and Lincoln 2010).

Biotech patents are an example of an area that saw sharp increases in ethnically Chinese names, followed by those of Indians and Japanese. While inventors with Chinese and Indian names are far more likely to be represented by US-owned firms, innovators with Japanese names work for Japanese companies. As such, patenting activity for those with Indian and Chinese names disproportionately benefits US companies. When examining cross-border collaborations with US inventors for patents filed by multiple inventors from different countries, partnerships with China, Japan, and India lag behind those with Germany, Canada, and the United Kingdom (Guang et al. 2024).

Firms that win the lottery to hire skilled H-1B immigrants tend to expand their skill-intensive activities, hire more overall, and innovate more (Kerr, Kerr, and Lincoln 2015). Using US Census data, Mahajan et al. (2024) find that hiring H-1B workers increases employment, revenues, and the survival of firms without significantly crowding out native college graduates. Lottery-winning firms even expand their hiring of native college graduates, the group that is the most substitutable to H-1B immigrants. On the other hand, Doran, Gelber, and Isen (2022) leverage a similar lottery but find only a modest increase in overall employment at these firms, and they argue that the use of H-1B visas does not significantly affect firms' patenting activities or innovation outcomes. The H-1B program plays an important role in new product development (Khanna and Lee 2018). H-1B migration (primarily from India) can explain 30–50 percent of aggregate productivity growth in US cities from 1990 to 2010, with modest positive wage effects for natives (Peri, Shih, and Sparber 2015). Such studies' findings may help explain why inflows of H-1B workers generally do not find large negative wage impacts: taken as a group, these immigrants are increasing overall demand for high-tech workers.

Asian-born immigrants have also played prominent roles in entrepreneurship, particularly Chinese and Indian immigrants in the tech sector. In the late 1990s, Chinese- and Indian-born engineers were running one out of every four US tech startups (Saxenian 1999), and a similar fraction had at least one immigrant founder (Wadhwa et al. 2007). More generally, immigrants are twice as likely to start new businesses as natives (Fairlie and Lofstrom 2015). In 2023, 44.8 percent of US Fortune 500 companies were founded by immigrants (from anywhere in the world) or their children (AIC 2023); in 2022, 55 percent of America's "unicorn" startups (privately held startups with total stock market capitalization over $1 billion) had an immigrant founder (Kapur and Vaishnav 2024). As of 2024, chief executive officers of giants like Google, Microsoft, Adobe, Cognizant, and IBM were born and educated in India. Indian-born CEOs of Fortune 500 companies are also found in other sectors of the economy, including technology (Micron Technology, Microchip Technology), consumer goods and retail (Starbucks, Chewy, Albertsons), technical services (FedEx, Cognizant, Honeywell, Jacobs Solutions), and pharmaceuticals (Vertex). Asian-born managers in tech companies often began their careers as programmers or scientists, then transitioned into managerial roles. Looking ahead at the race for artificial intelligence startups, a recent analysis found that nearly two-thirds of the Forbes top 50 artificial intelligence startups had an immigrant founder, with Indian entrepreneurs leading the pack with 10 startups (Anderson 2023).

Why are immigrants from Asia disproportionately involved in innovation? Some open questions remain here. One possibility is that education systems in their home countries might shift the distribution of innovative ability among migrants (Martellini, Schoellman, and Sockin 2024). Perhaps Asian-born students and workers have been trained in skills that complement those of their US-educated counterparts, and working in teams with complementary skills facilitates innovative activity. Perhaps emigration involves self-selection on risk-taking, which may correlate with high entrepreneurial ambitions. Another possibility is that the sheer size of Asian populations provides a larger mass in the right tail of the talent distribution. Furthermore, while the H-1B visa is not specifically designed for entrepreneurs, it does prioritize high-skill specialty occupations that may lead to increased product innovation and technological application (Hunt 2011).

However, these theories do not explain why such innovation is concentrated in the US economy, rather than back home in Asia. Therefore, it is plausible that barriers to entrepreneurship and innovation exist in countries like India and China, which may include bureaucratic hurdles in starting a business, complex labor laws, inadequate contract enforcement, and land acquisition restrictions. Indeed, Kapur and McHale (2005) argue that Indian talent emigrates not just for the economic opportunities and research infrastructure in the US economy, but also because of bureaucratic and domestic constraints at home.

Over time, immigrant innovation and entrepreneurship have also affected countries in Asia through knowledge transfers from abroad back to these countries (Kerr 2008). This feedback can be particularly pronounced within multinational enterprises that facilitate diffusion of research and development to India

(Choudhury 2016), business practices to China (Giannetti, Liao, and Yu 2015), and trade links to Vietnam (Parsons and Vézina 2018). Links with a preexisting diaspora, combined with return migration and remittances, also help facilitate the growth of an entrepreneurial tech sector in countries of Asian origin (Khanna and Morales 2025; Giannetti, Liao, and Yu 2015; Saxenian 1999). As such, the migration of entrepreneurs may also eventually lead to business opportunities in Asia, in the form of business offshoring and supply-chain linkages (Parsons and Vézina 2018).[4] As mentioned earlier, return migration rates are higher for Chinese, Korean, and Japanese migrants than for migrants from India. This pattern may reflect several factors, including the relatively more attractive income levels and opportunities available back home in China, Korea, and Japan. For migrants from India, a combination of comfort with the English language and the Indian diaspora within firms allows Indians to stay in the United States. Finally, small families in Japan, Korea, and China may also influence return rates, to the extent that return migration reflects family obligations and the need for old-age support for parents.

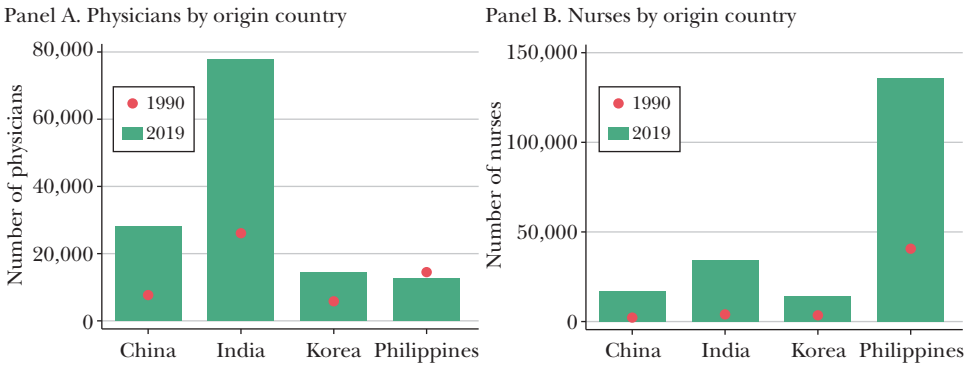
Medical Professionals: Nurses and Physicians

Asian-born immigrants are highly represented in the US labor market as nurses and physicians, as shown in Figure 5. In 2023, according to the Association of American Medical Colleges (2024), one in four active US physicians were international medical graduates. In 2016, migrants from Asia made up 17.7 percent of all US physicians and surgeons (Patel et al. 2018). Physicians are more likely to be from India, with recent growth in the number of physicians from China. The Philippines, on the other hand, is an important source of the nursing workforce in the United States. Just before the Covid-19 pandemic, there were 136,000 Filipino nurses and 78,000 Indian-born physicians in the United States. Migrants from Asia—and in particular, China, India, and the Philippines—represent 20 percent of the growth in the US physician population and 11 percent of the increase in the US nurse population between 1990 and 2019.

Again, as in the earlier discussion, both demand and supply factors play a role. US medical schools and residency programs have not expanded quickly enough to meet the rising patient demand, partly due to funding caps and constraints imposed by medical schools. As a result, the AAMC (2021) expects a shortfall of 124,000 physicians by 2034, particularly as the US-born population ages rapidly. These limited medical school slots have led to a physician shortfall, boosting demand for international medical residents. Patients treated by foreign-trained doctors in US hospitals had similar or slightly lower mortality than those treated by US-trained doctors, suggesting high quality of care (Tsugawa et al. 2017).

[4] This body of work combines matched firm-level data with regional migration records to study knowledge diffusion. By leveraging shocks to return migration, multinational expansion, and diaspora linkages, these studies identify how immigrant scientists and entrepreneurs transmit know-how and business practices back to Asia.

Figure 5
Physicians and Nurses in the United States, 1990 and 2019



Source: Author's calculations using data from the US Census Bureau.
Note: Figure shows the number of physicians (left panel) and number of nurses (right panel) working in the United States by country of origin for four major Asian countries. The red dots are total number of workers from the 1990 US Census, and the green bars represent numbers from the 2019 ACS.

The US Department of Labor designates nursing as a Schedule A occupation, signalling that this occupation has a shortage of workers and thus streamlining the immigration process. The Philippines, in particular, has been training a surplus of nurses for export and has been sending nurses to the United States since the 1960s. Filipino nurses help the United States meet nursing demand and have comparable performance to US-trained nurses (Aiken 2007; Cortés and Pan 2015).

The effect of Asian migrants working as doctors and nurses is felt with particular force in medically underserved areas, in part because of rules in US visa programs. Medical residents arrive on short-term J-1 visas but must return home unless they work in medically underserved or "health professional shortage areas." The Conrad-30 visa program helps locate immigrants to these areas, often in rural counties, where there are shortages of both primary care and specialty care physicians. The visa waivers under the Conrad-30 are concentrated among those from Asia, with the majority of participants coming from the Philippines, India, Nigeria, and Pakistan (Ranasinghe 2015; Crouse and Munson 2006).

Even over the long run, immigrants are more likely to stay in rural areas (Goodfellow et al. 2016) and provide relatively high-quality care (Tsugawa et al. 2017). Cortés and Pan (2015) examine the performance of foreign-educated nurses in the United States, using administrative data on licensing exams, hospital employment, and patient outcomes. They find that foreign-educated nurses (mostly from the Philippines) are positively selected, playing an important role in alleviating nurse shortages without substantially reducing care quality.

As a result, in US rural areas with the lowest income and education levels, international medical graduates often comprise more than 30 percent of the physician workforce (AIC 2018). Braga, Khanna, and Turner (2024) leverage changes to

the Conrad-30 program state-level cap to show how Asian-born physicians locate in shortage areas. They highlight the critical role Asian-born physicians played during the Covid-19 pandemic, particularly in medically underserved areas. Given the shortage of US-born health care workers, "health professional shortage areas," Asian immigrants likely improve health outcomes without "crowding out" natives. More broadly, the 17 percent of US healthcare professionals who are foreign-born are more likely to work longer hours, cover night shifts, and serve in nursing homes and underserved areas than their US-born counterparts (Commodore-Mensah et al. 2021).

From the supply side, the high representation of the Philippines in nursing and India in physician labor supply reflects the potential for training in their countries of origin. Each of these countries has high-quality nursing and medical schools that select and train a large number of physicians and nurses. The Philippines has a US-modeled nursing education system, and both countries use English-language instruction, facilitating easy transitions into American healthcare. Indeed, physicians and nurses from these countries are also likely to migrate to other parts of the world, including the United Kingdom and parts of East Asia. The Philippines has become the world's largest exporter of nurses, accounting for roughly 25 percent of all overseas nurses globally. India hosts the largest number of medical schools globally and is the leading source of internationally trained physicians, sending doctors to the United States, the United Kingdom, Canada, and Australia. This expansion continues to occur in India, with a near doubling in both the number of new medical colleges and enrollment between 2013 and 2024. India also ranks second only to the Philippines in training nurses, with nearly 56,000 Indian-trained nurses employed in those same four countries (Carzaniga et al. 2019). This is equivalent to around 3 percent of all registered nurses in India. Given low wages at home and high burnout rates in origin country healthcare sectors (Alibudbud 2023), many choose to migrate abroad.

One natural concern is that sending nurses or doctors abroad might leave countries like India or the Philippines with fewer healthcare workers for their own populations. More than half of Filipino nurses work overseas, and the World Health Organization actually predicts a shortfall of nurses within the Philippines (Elmaco 2022). But evidence on the effects of migration is more nuanced, because incentives for potential migration generate human capital accumulation back home. Abarcar and Theoharides (2024) exploit changes in US visa quotas to show that as the prospects for nurse migration to the United States improved, Filipino students increasingly acquired skills valued abroad. Among those who migrated, remittances substantially bolstered local economies, while many who remained in the Philippines continued to work in domestic healthcare. These remittances often financed further investments in health and education within migrant-sending communities. Khanna et al. (2022) similarly exploit exchange-rate shocks to show that remittance inflows to the Philippines spurred additional skill accumulation and human-capital investment. Similarly, while thousands of doctors have migrated from India, the country has simultaneously expanded its medical college capacity dramatically over the last few decades—in part due to the prestige and opportunities associated with the global medical labor market. Thus, the narrative of "brain drain" and "brain

gain" from training and migration involves possible trade-offs and feedback loops. Origin countries benefit from remittances and the incentive for more medical training; however, they also lose practitioners to migration.

For the United States, population aging and persistent domestic training bottlenecks make it likely that the United States will continue to have a high demand for foreign-educated health workers. Asian-born nurses and physicians, in particular, are poised to remain vital in meeting America's healthcare needs, helping to alleviate staffing shortfalls that would be far more severe in their absence.

Looking Ahead

Looking ahead, the US economy has become part of a larger global competition for talent, with India and China playing a pivotal role in international migration related to science, technology, engineering, and mathematics (Kerr et al. 2016). The United States still enjoys an unmatched dominance in technology and higher education. But as the United States debates additional immigration restrictions, Asian flows of students and skilled workers may shift to Canada, the United Kingdom, countries in Europe, and Australia. For instance, Canada offers work permits and paths to residency for foreign graduates, and the United Kingdom recently reinstated a two-year post-study work visa. Canada's Global Talent Stream has successfully attracted tens of thousands of tech workers. China has heavily invested in its own higher education sector, building on return migrants from the United States. Of course, competitor countries face their own political headwinds. For instance, the largest competitor countries have all discussed restrictions to student visas in 2024. While the Australian government is still debating caps (Cassidy 2024), UK universities have agreed to restrict foreign students if they can raise tuition rates (Adams 2024), and Canada has uniquely started explicitly capping international student permits (Lone 2025).

Some of the best-performing US sectors in the global economy are linked to skilled migration from Asia. For example, in 2024, computer and information services saw an export of $79.2 billion, and higher education an export of $56 billion (Bureau of Economic Analysis 2025). Looking ahead to what may be a coming boom in artificial intelligence, one-fifth of the top artificial intelligence startups were started by Indian migrants in the United States (Anderson 2023).

How can the United States remain a magnet for Asian-born talent? Immigration policy will play an important role. Indians and Chinese immigrants face long wait times (sometimes more than a decade) for permanent residency status because of country-specific quotas for green cards.[5] Expanding country quotas may allow the United States to attract and hold on to top talent from India and China. The

[5] Employment-based green cards—that is, "EB cards"—have five categories. In 2025, EB-5 cases are being processed for Indians who applied in 2019 and for Chinese nationals who applied in 2014. EB-2 and EB-3 cases are currently under consideration for Indian applicants who submitted their applications in 2013 and Chinese nationals who submitted their applications in 2020.

H-1B visa is consistently oversubscribed, and the US Congress has the ability to raise the cap. H-1B reform, specifically addressing how to select the most productive workers (as opposed to a lottery), is a much-discussed issue among policymakers. How the United States navigates these issues will shape not only its own economic future, but also the trajectory of innovation globally and the development paths of Asian countries.

■ *Thank you to Sam Bazzi, Breno Braga, and Hannah Postel for detailed discussions and to Braden Clark for excellent research assistance.*

References

Abarcar, Paolo, and Caroline Theoharides. 2024. "Medical Worker Migration and Origin-Country Human Capital: Evidence from US Visa Policy." *Review of Economics and Statistics* 106 (1): 20–35.

Adams, Richard. 2024. "English Universities Offer to Curb Foreign Students If They Can Raise Tuition Fees." *The Guardian*, September 17. https://www.theguardian.com/education/2024/sep/17/english-universities-offer-to-curb-foreign-students-if-they-can-raise-tuition-fees.

ACS (American Community Survey). 2019. *US Census Bureau, 1990–2019, Census and American Community Survey 1-Year Estimates*. https://www.census.gov/programs-surveys/acs/data.html (accessed on February 14, 2025).

Alibudbud, Rowalt. 2023. "Addressing the Burnout and Shortage of Nurses in the Philippines." *SAGE Open Nursing* 9.

AIC (American Immigration Council). 2018. *Foreign-Trained Doctors Are Critical to Serving Many US Communities*. American Immigration Council.

AIC (American Immigration Council). 2022. "Foreign-Born STEM Workers in the United States." American Immigration Council. https://www.americanimmigrationcouncil.org/sites/default/files/research/foreign-born_stem_workers_in_the_united_states_final_0.pdf.

AIC (American Immigration Council). 2023. *New American Fortune 500 in 2023*. American Immigration Council.

Anderson, Stuart. 2023. *AI and Immigrants*. National Foundation for American Policy.

Anelli, Massimo, Kevin Shih, and Kevin Williams. 2023. "Foreign Students in College and the Supply of STEM Graduates." *Journal of Labor Economics* 41 (2): 511–63.

Arora, Ashish, and Suma Athreye. 2002. "The Software Industry and India's Economic Development." *Information Economics and Policy* 14 (2): 253–73.

AAMC (Association of American Medical Colleges). 2021. *The Complexities of Physician Supply and Demand: Projections from 2019 to 2034*. Association of American Medical.

AAMC (Association of American Medical Colleges). 2024. *2024 Key Findings and Definitions*. US Physician Workforce Data Dashboard. https://www.aamc.org/data-reports/data/2024-key-findings-and-definitions (accessed April 19, 2025).

Bhatnagar, Subhash. 2006. "India's Software Industry." In *Technology, Adaptation, and Exports: How Some Developing Countries Got It Right*, edited by Vandana Chandra, 49–82. World Bank.

Black, Grant C., and Paula E. Stephan. 2010. "The Economics of University Science and the Role of Foreign Graduate Students and Postdoctoral Scholars." In *American Universities in a Global Market*, edited by Charles T. Clotfelter, 129–61. University of Chicago Press.

Bound, John, Breno Braga, Joseph M. Golden, and Gaurav Khanna. 2015a. "Recruitment of Foreigners

in the Market for Computer Scientists in the United States." *Journal of Labor Economics* 33 (S1): S187–223.

Bound, John, Breno Braga, Gaurav Khanna, and Sarah Turner. 2020. "A Passage to America: University Funding and International Students." *American Economic Journal: Economic Policy* 12 (1): 97–126.

Bound, John, Breno Braga, Gaurav Khanna, and Sarah Turner. 2021. "The Globalization of Postsecondary Education: The Role of International Students in the US Higher Education System." *Journal of Economic Perspectives* 35 (1): 163–84.

Bound, John, Murat Demirci, Gaurav Khanna, and Sarah Turner. 2015b. "Finishing Degrees and Finding Jobs: US Higher Education and the Flow of Foreign IT Workers." *Innovation Policy and the Economy* 15: 27–72.

Braga, Breno, Gaurav Khanna, and Sarah Turner. 2024. "Migration Policy and the Supply of Foreign Physicians: Evidence from the Conrad 30 Waiver Program." *Journal of Economic Behavior and Organization* 226: 106682.

Bureau of Economic Analysis. 2025. *International Transactions, International Services, and International Investment Position Tables*. BEA International Data. https://apps.bea.gov/iTable/?reqid=62&step=6&isuri=1&tablelist=51&product=1 (accessed on March 1, 2025).

Bureau of Labor Statistics. 2025. *Consumer Price Index (CPI) by Category*. US Department of Labor. https://www.bls.gov/cpi/ (accessed April 16, 2025).

Carzaniga, Antonia, Ibadat Dhillon, Joscelyn Magdeleine, and Lihui Xu. 2019. "International Health Worker Mobility and Trade in Services." WTO Staff Working Paper ERSD-2019-13.

Cassidy, Caitlin. 2024. "Australian Government Moves to Slow Foreign Student Visas after Failure of Cap Plan." *The Guardian*, December 18. https://www.theguardian.com/australia-news/2024/dec/19/labor-australia-foreign-student-visa-cap-plan-new-policy-processing.

Chang, Keng-Chi, Ruixue Jia, Steven Liao, and Margaret E. Roberts. 2025. "The Relocation of Students from China in U.S. Higher Education." Unpublished.

Chellaraj, Gnanaraj, Keith E. Maskus, and Aaditya Mattoo. 2008. "The Contribution of Skilled Immigration and International Graduate Students to US Innovation." *Review of International Economics* 16 (3): 444–62.

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

Clemens, Michael A. 2013. "Why Do Programmers Earn More in Houston than Hyderabad? Evidence from Randomized Processing of US Visas." *American Economic Review* 103 (3): 198–202.

Commodore-Mensah, Yvonne, Kelli DePriest, Laura J. Samuel, Ginger Hanson, Rita D'Aoust, and Eric P. Slade. 2021. "Prevalence and Characteristics of Non–US-Born and US-Born Health Care Professionals, 2010–2018." *JAMA Network Open* 4 (4): e218396.

Cortés, Patricia, and Jessica Pan. 2015. "The Relative Quality of Foreign-Educated Nurses in the United States." *Journal of Human Resources* 50 (4): 1009–50.

Crouse, Byron J., and Randy L. Munson. 2006. "The Effect of the Physician J-1 Visa Waiver on Rural Wisconsin." *Wisconsin Medical Journal* 105 (7): 16–20.

De Nardi, Mariacristina, Eric French, John Bailey Jones, and Jeremy McCauley. 2015. "Medical Spending of the US Elderly." *Fiscal Studies* 37 (3-4): 717–47.

Doran, Kirk, Alexander Gelber, and Adam Isen. 2022. "The Effects of High-Skilled Immigration Policy on Firms: Evidence from Visa Lotteries." *Journal of Political Economy* 130 (10): 2501–33.

Elmaco, Jenny Lind. 2022. "Philippine Nurse Migration: Assessing Vulnerabilities and Accessing Opportunities during the COVID-19 Pandemic." In *Governing Migration for Development from the Global South: Challenges and Opportunities*, Vol. 14, *International Development Policy*, edited by Dêlidji Eric Degila and Valeria Marina Valle, 208–33. Brill Nijhoff.

Fairlie, Robert W., and Magnus Lofstrom. 2015. "Immigration and Entrepreneurship." In Handbook of the Economics of International Migration, Vol. 1, edited by Barry R. Chiswick and Paul W. Miller, 877–911. North-Holland.

Finn, Michael G., and Leigh Ann Pennington. 2018. *Stay Rates of Foreign Doctorate Recipients from US Universities, 2013*. Oak Ridge Institute for Science and Education.

Flood, S., M. King, R. Rodgers, S. Ruggles, and J. R. Warren. 2024. *Integrated Public Use Microdata Series, Current Population Survey: Version 12.0 [dataset]*. Minneapolis, MN: Minnesota Population Center. Retrieved December 2024 from https://cps.ipums.org/cps/.

Gaulé, Patrick, and Mario Piacentini. 2013. "Chinese Graduate Students and US Scientific Productivity."

Review of Economics and Statistics 95 (2): 698–701.

Giannetti, Mariassunta, Guanmin Liao, and Xiaoyun Yu. 2015. "The Brain Gain of Corporate Boards: Evidence from China." *Journal of Finance* 70 (4): 1629–82.

Goodfellow, Amelia, Jesus G. Ulloa, Patrick T. Dowling, Efrain Talamantes, Somil Chheda, Curtis Bone, and Gerardo Moreno. 2016. "Predictors of Primary Care Physician Practice Location in Underserved Urban or Rural Areas in the United States: A Systematic Literature Review." *Academic Medicine* 91 (9): 1313–21.

Guang, Lei, Ruixue Jia, Jingwen Liu, and Young Yang. 2024. "Chinese Talent, American Enterprise: Five Takeaways of How Chinese Talent Contributes to Biotech Innovation in the US." China Data Lab (blog), April. https://chinadatalab.ucsd.edu/viz-blog/chinese-talent-american-enterprise-five-takeaways-of-how-chinese-talent-contributes-to-biotech-innovation-in-the-u-s/.

Hunt, Jennifer. 2011. "Which Immigrants Are Most Innovative and Entrepreneurial? Distinctions by Entry Visa." *Journal of Labor Economics* 29 (3): 417–57.

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle. 2010. "How Much Does Immigration Boost Innovation?" *American Economic Journal: Macroeconomics* 2 (2): 31–56.

ICE (US Immigration and Customs Enforcement). 2016. "SEVP Policy Guidance S13.2: The Form I-20 and the English Proficiency Field." https://www.ice.gov/sites/default/files/documents/Document/2016/sevp-PGS132-i20.pdf.

Institute of International Education. 2024. *Leading Places of Origin of International Scholars, 2023/24.* Open Doors. https://opendoorsdata.org/data/international-scholars/leading-places-of-origin/ (accessed on April 15, 2025).

Jia, Ruixue, Gaurav Khanna, Hongbin Li, and Yuli Xu. 2025. "The Ripple Effects of China's College Expansion on American Universities." NBER Working Paper 34391.

Jorgenson, Dale W., Mun S. Ho, and Jon D. Samuels. 2016. "The Impact of Information Technology on Postwar US Economic Growth." *Telecommunications Policy* 40 (5): 398–411.

Jung, Juergen, Chung Tran, and Matthew Chambers. 2017. "Aging and Health Financing in the US: A General Equilibrium Analysis." *European Economic Review* 100: 428–62.

Kapur, Devesh, and John McHale. 2005. *Give Us Your Best and Brightest: The Global Hunt for Talent and Its Impact on the Developing World.* Center for Global Development and Brookings Institution Press.

Kapur, Devesh and Milan Vaishnav. 2024. "Industrial Policy Needs an Immigration Policy: Why Bringing Jobs Back to the United States Requires Letting in More Foreign Workers." Foreign Affairs, August 22. https://www.foreignaffairs.com/united-states/industrial-policy-needs-immigration-policy.

Kerr, Sari Pekkala, William R. Kerr, and William F. Lincoln. 2015. "Skilled Immigration and the Employment Structures of US Firms." *Journal of Labor Economics* 33 (S1): S147–86.

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

Kerr, William R. 2008. "Ethnic Scientific Communities and International Technology Diffusion." *Review of Economics and Statistics* 90 (3): 518–37.

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

Khanna, Gaurav. 2026. *Data and Code for: "From Asia, with Skills."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E240483V1.

Khanna, Gaurav, and Munseob Lee. 2018. "High-Skill Immigration, Innovation, and Creative Destruction." NBER Working Paper 24824.

Khanna, Gaurav, and Nicolas Morales. 2025. "The IT Boom and Other Unintended Consequences of Chasing the American Dream." Center for Global Development Working Paper 460.

Khanna, Gaurav, Emir Murathanoglu, Caroline B. Theoharides, and Dean Yang. 2022. "Abundance from Abroad: Migrant Income and Long-Run Economic Development." NBER Working Paper 29862.

Khanna, Gaurav, Kevin Shih, Ariel Weinberger, Mingzhi Xu, and Miaojie Yu. 2023. "Trade Liberalization and Chinese Students in US Higher Education." *Review of Economics and Statistics* 107 (5): 1291–1309.

Lee, Saerom (Ronnie), and Britta Glennon. 2023. "The Effect of Immigration Policy on Founding Location Choice: Evidence from Canada's Start-Up Visa Program." NBER Working Paper 31634.

Leiner, Barry M., Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Lawrence G. Roberts, and Stephen S. Wolff. 1997. "The Past and Future History of the Internet." *Communications of the ACM* 40 (2): 102–08.

Lone, Wa. 2025 "Canada Reduces International Student Permits for Second Year." *Reuters*, January 24.

https://www.reuters.com/world/americas/canada-reduces-international-student-permits-second-year-2025-01-24/.

Mahajan, Parag, Nicolas Morales, Kevin Shih, Mingyu Chen, and Agostina Brinatti. 2024. "The Impact of Immigration on Firms and Workers: Insights from the H-1B Lottery." IZA Discussion Paper 16917.

Martellini, Paolo, Todd Schoellman, and Jason Sockin. 2024. "The Global Distribution of College Graduate Quality." *Journal of Political Economy* 132 (2): 434–83.

Naplitano, Janet (2016). *Response to the Audit: The University of California: Its Admissions and Financial Decisions Have Disadvantaged California Resident Students.* Public letter to California State Auditor), March 8, 2016, University of California.

NCSES (National Center for Science, and Engineering Statistics). 2025. *Most US-Trained Science and Engineering Doctorate Recipients on Temporary Visas Remain in the United States.* Technical Report NSF 25-325. US National Science Foundation.

NSF (National Science Foundation). 2019. *Science and Engineering Indicators.* https://ncses.nsf.gov/indicators (accessed June 2019).

Orr, Robert. 2020. "The Planning of US Physician Shortages." Niskanen Center, September 8. https://www.niskanencenter.org/the-planning-of-u-s-physician-shortages/.

Ottaviano, Gianmarco I. P., and Giovanni Peri. 2012. "Rethinking the Effect of Immigration on Wages." *Journal of the European Economic Association* 10 (1): 152–97.

Parsons, Christopher, and Pierre-Louis Vézina. 2018. "Migrant Networks and Trade: The Vietnamese Boat People as a Natural Experiment." *Economic Journal* 128 (612): F210–34.

Patel, Yash M., Dan P. Ly, Tanner Hicks, and Anupam B. Jena. 2018. "Proportion of Non–US Born and Noncitizen Health Care Professionals in the United States in 2016." *JAMA* 320 (21): 2265–67.

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

Ranasinghe, Padmini D. 2015. "International Medical Graduates in the US Physician Workforce." *Journal of Osteopathic Medicine* 115 (4): 236–41.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3 (2): 135–46.

Saxenian, AnnaLee. 1999. "Silicon Valley's New Immigrant Entrepreneurs." University of California-San Diego Center for Comparative Immigration Studies Working Paper 15.

Shih, Kevin. 2017. "Do International Students Crowd-Out or Cross-Subsidize Americans in Higher Education?" *Journal of Public Economics* 156: 170–84.

Smith, Bethany, Caren A. Arbeit, Herbert Thompson, Michael I. Yamaner, and NCSES. 2024. *Graduate Enrollment and Postdoctoral Appointments in Science, Engineering, and Health Rise, Driven Largely by Increases in the Number of Women and Temporary Visa Holders.* Technical Report NSF 25-316. US National Science Foundation.

Stuen, Eric T., Ahmed Mushfiq Mobarak, and Keith E. Maskus. 2012. "Skilled Immigration and Innovation: Evidence from Enrolment Fluctuations in US Doctoral Programmes." *Economic Journal* 122 (565): 1143–76.

Tsugawa, Yusuke, Anupam B. Jena, E. John Orav, and Ashish K. Jha. 2017. "Quality of Care Delivered by General Internists in US Hospitals Who Graduated from Foreign versus US Medical Schools: Observational Study." *British Medical Journal* 356: j273.

Turner, Patrick S. 2022. "High-Skilled Immigration and the Labor Market: Evidence from the H-1B Visa Program." *Journal of Policy Analysis and Management* 41 (1): 92–130.

UNESCO. 2025. *Global Flow of Tertiary-Level Students, Institute for Statistics.* https://uis.unesco.org/en/uis-student-flow/ (accessed April 15, 2025).

US Citizenship and Immigration Services (USCIS). 2017. *F-1 Visa Records.* Freedom of Information Act (FOIA) Request, received from the Department of Homeland Security, Washington, DC.

US Department of State. 2025. *Nonimmigrant Visa Statistics.* https://travel.state.gov/content/travel/en/legal/visa-law0/visa-statistics/nonimmigrant-visa-statistics.html (accessed April 23, 2025).

Wadhwa, Vivek, AnnaLee Saxenian, Ben A. Rissing, and Gary Gereffi. 2007. "America's New Immigrant Entrepreneurs: Part 1." Duke Science, Technology, and Innovation Paper 23.

World Bank. 2025. *World Development Indicators.* https://databank.worldbank.org/source/world-development-indicators (accessed March 31, 2025).

Journal of Economic Perspectives—Volume 40, Number 1—Winter 2026—Pages 241–248

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by e-mail at <taylort@macalester.edu>, or c/o Journal of Economic Perspectives, Macalester College, 1600 Grand Ave., Saint Paul, MN 55105.

Smorgasbord

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2025 for "having explained innovation-driven economic growth" was awarded to Joel Mokyr, Philippe Aghion, and Peter Howitt (https://www.nobelprize.org/prizes/economic-sciences/2025). From the "Popular Information" description: "Through his research in economic history, Joel Mokyr has demonstrated that a continual flow of *useful knowledge* is necessary. This useful knowledge has two parts: the first is what Mokyr refers to as *propositional knowledge*, a systematic description of regularities in the natural world that demonstrate *why* something works; the second is *prescriptive knowledge*, such as practical instructions, drawings or recipes that describe what is necessary for something to work. . . . The 16th and 17th centuries witnessed the Scientific Revolution as part of the Enlightenment. Scientists began to insist upon

■ *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.20251497.

precise measurement methods, controlled experiments, and that results should be reproducible, leading to improved feedback between propositional and prescriptive knowledge. This increased the accumulation of useful knowledge that could be utilised in the production of goods and services. . . . [I]nspired by modern data, Philippe Aghion and Peter Howitt constructed a mathematical economic model that . . . can be used to analyse whether there is an optimal volume of R&D . . . It turned out that the answer was far from simple, because two mechanisms pull in different directions. The first mechanism is based upon companies that invest in R&D understanding that their current profits from an innovation will not continue forever. . . . From the perspective of society, however, the value of the old innovation does not disappear, because the new one builds upon the old knowledge. Outcompeted innovations thus have a greater value for society than for the companies that develop them, which makes the private incentives for R&D smaller than the gains to society as a whole. Society can therefore benefit from subsidising R&D. The second mechanism looks at how, when one company succeeds in pushing another from the top of the ladder . . . Therefore, even if the new innovation is only slightly better than the old one, profits may be significant and larger than the socioeconomic gains. Therefore, from a socioeconomic perspective, investments in R&D can be too large; technological development can be too rapid and growth too high. This creates arguments against society subsidising R&D. Which of these two forces dominates depends on a range of factors, which vary from market to market and time to time."

A group of IMF economists draw lessons from a database of fiscal rules in more than 120 countries in "Fiscal Guardrails against High Debt and Looming Spending Pressures" (IMF Staff Discussion Note SDN/2025/004, September 2025, by Julien Acalin, Virginia Alonso-Albarran, Clara Arroyo, Waikei Raphael Lam, Leonardo Martinez, Anh D. M. Nguyen, Francisco Roch, Galen Sher, and Alexandra Solovyeva, https://www.imf.org/en/publications/staff-discussion-notes/issues/2025/09/22/fiscal-guardrails-against-high-debt-and-looming-spending-pressures-569841). "Although earlier fiscal rules were often too rigid, efforts to introduce greater flexibility have not translated into stronger compliance. . . . [F]ewer than two-thirds of countries adhere to their deficit rules on average, with lower share for emerging market and developing countries and debt rules. . . . Fiscal deficits four years after the pandemic continue to exceed fiscal rule limits by a median of 2.0–2.5 percentage points of GDP for about 40 percent of advanced economies and 60 percent of EMDEs [emerging market and developing economies]. In most countries, public debt has surpassed the ceilings in the debt rule by an average of 25 percentage points of GDP. Such large deviations from fiscal rule limits in many countries are driven by both severe shocks and limitations in the design of fiscal rules. During the severe shocks, the magnitudes and the share of countries that deviate from fiscal rule limits increased as expenditures or deficits tend to rise. But even in normal times, some countries have deficits and debt persistently exceeding their fiscal rule limits, partly because of multiple exclusions from the rules, limited fiscal oversight, or lack of fiscal adjustments to reduce debt and deficits. In recent years, fiscal adjustments have been limited, complicating the return to fiscal rule limits."

Andrew T. Levin and Christina Parajon Skinner discuss "Central Bank Undersight: Assessing the Fed's Accountability to Congress" (*Vanderbilt Law Review*, 2024, 77:6, pp. 1769–830, https://scholarship.law.vanderbilt.edu/vlr/vol77/iss6/3/). "Over the past 15 years, however, the scope and complexity of monetary policy has outpaced Congress's ability to monitor these policies through existing mechanisms of oversight. For example, internal shifts in the Fed's governance and power dynamics have led to the disappearance of dissents on monetary policy decisions, thereby hampering legislators' abilities to discern the range of views that have informed those decisions. Moreover, in conducting its latest round of securities purchases ('QE4') during 2020–22, the Fed did not provide legislators with cost-benefit analyses or risk assessments at any stage of the program. Indeed, QE4 is now likely to cost taxpayers more than $1 trillion, but its efficacy has still not been scrutinized by any external reviews."

Gabriel E. Lade and Aaron Smith provide an overview of "Biofuels: Past, Present, and Future" (*Annual Review of Resource Economics*, 2025, 17: pp. 10–25, https://www.annualreviews.org/content/journals/10.1146/annurev-resource-011724-082950). "Before 2005, biofuels comprised little of the country's energy consumption; ethanol made up less than 3% of gasoline, and biomass-based diesel use was negligible. . . . The Energy Policy Act and subsequent RFS2 mandates . . . transformed the US agricultural sector and gasoline markets; around one-third of all corn produced in the United States is used for fuel production, and nearly every gallon of gasoline sold in the United States contains at least 10% corn ethanol. . . . Although biomass-based diesel markets are much smaller than ethanol markets by comparison, their growth, paired with the recent surge in renewable diesel, means these fuels now demand more than 40% of the soybean oil produced in the United States. The past and future benefits of these policies beyond increasing crop demand are less clear. Expanding cropland generates substantial carbon emissions that can offset the gains from burning less fossil fuel. Low-carbon biofuels are limited by high cost and the low supply of non-crop feedstocks such as animal fats and used cooking oil. Scaling production technologies for liquid cellulosic ethanol proved much more challenging and costly than thought in 2007. . . . In summary, while the fuels and focus are different than in 2007, the industry in 2025 has largely returned to past controversies about land use change, food prices, and the high cost of low-carbon feedstocks."

Gabriel Felbermayr, T. Clifton Morgan, Constantinos Syropoulos, and Yoto V. Yotov review "Economic Sanctions: Stylized Facts and Quantitative Evidence" (*Annual Review of Economics*, 2025, 17: pp. 175–95, https://www.annualreviews.org/content/journals/10.1146/annurev-economics-081623-020909). "According to the newest version of the Global Sanctions Data Base, the number of sanction programs in place globally has shot up from about 200 10 years ago to about 600 in 2023. What is more, about 12% of all existing country pairs and 27% of world trade are currently affected by some type of sanction. In their various forms, sanctions are the leading geoeconomic tool aiming to coerce foreign governments into actions that they would not undertake otherwise. . . . [S]anction processes are much better

understood today than they were 30 years ago. The political science community has come to accept what economists already knew (i. e., that sanctions bring substantial economic effects), and economists have come to accept what political scientists have long understood (i. e., that substantial economic costs do not always bring changes in policy). Over that same period of time, the use of sanctions has dramatically increased, and they have come to affect many more bilateral economic relationships. This is a puzzling phenomenon: If sanctions are costly and frequently fail to deliver the desired policy objectives, why have they become so endemic?"

Robin Greenwood, Robert Ialenti, and David Scharfstein explore "The Evolution of Financial Services in the United States" (*Annual Review of Financial Economics*, 2025, 17: pp. 189–206, https://www.annualreviews.org/content/journals/10.1146/annurev-financial-082123-105625). From the abstract: "This article surveys the literature on the historical growth and transformation of the US financial sector. The sector expanded rapidly between 1980 and 2006, during which its contribution to GDP rose from 4.8% to 7.6%. After the global financial crisis, the size of the sector stabilized at approximately 7% of GDP. After reviewing this literature, we examine recent developments, including the continued growth of high-fee alternative asset management and the shift away from banks to lending by nonbank financial intermediaries. We interpret both the growth and recent evolution of the sector as reflecting a continued transition to a more market-based financial system, with risk migrating away from banks and into markets."

The *OECD International Migration Outlook 2025*, along with its usual overview of trends in migration, migration policy, and migrant integration policy, includes a chapter on "International migration of health professionals to OECD countries," written by Ave Lauren, José Ramalho, Jean-Christophe Dumont, Gaetan Lafortune, Agya Mahat, and Tapas Nair (https://www.oecd.org/en/publications/2025/11/international-migration-outlook-2025_355ae9fd/full-report/international-migration-of-health-professionals-to-oecd-countries_fea88ae4.html). "Over the past two decades, the overall share of foreign-born health professionals in OECD countries has increased steadily. In countries with consistent data over the period, the total number of foreign-born doctors rose by 86% between 2000/01 and 2020/21, while the number of foreign-born nurses grew by nearly two and a half times . . . In both cases, this growth outpaced the general increase in the total number of doctors and nurses, which rose by 41% and 48%, respectively. . . . Among the main countries of residence, Germany and Australia saw the number of foreign-born doctors nearly triple. The United Kingdom experienced a doubling, and more moderate increases were observed in the United States and France. A similar pattern is evident among foreign-born nurses. Finland saw the steepest rise, with numbers increasing almost eightfold though starting from a very low level in 2000/01. In Norway, they increased more than fourfold. In Germany, Ireland and New Zealand, the numbers more than tripled, while in Australia and Spain, they nearly tripled. Switzerland also recorded a significant increase. Among the other major countries of residence, Canada, the United Kingdom and the United States all saw their numbers more than double . . . In absolute terms, the United States remains the primary country

of residence for both foreign-born doctors and nurses. . . . Among all foreign-born health professionals in OECD countries, 36% of all foreign-born doctors and 42% of nurses were practising in the United States in 2020/21."

Brian Albrecht, Geoffrey A. Manne, David Teece, and Mario A. Zúñiga discuss "From Moore's Law to Market Rivalry: The Economic Forces That Shape the Semiconductor Manufacturing Industry" (International Center for Law & Economics, November 12, 2025, https://laweconcenter.org/resources/from-moores-law-to-market-rivalry-the-economic-forces-that-shape-the-semiconductor-manufacturing-industry/). "Two unique, interconnected technological imperatives define semiconductor manufacturing competition. Moore's Law began as a rule of thumb about transistor counts doubling at a steady cadence. Today its practical meaning is broader: customers expect predictable gains in overall performance per watt and per dollar, and delivered on schedule. That pushes the industry through a relentless metronome of "beats," where each node (manufacturing generation) must arrive on time with real, measurable gains. Falling behind by a single beat can cost customers for years. Rock's Law is the darker side of Moore's Law. It holds that, as chips get denser, the cost of making them rises exponentially. Advanced fabs now cost $10–20 billion, and a single high-NA EUV scanner (the heart of leading-edge lithography) runs north of $400 million. These are not scale or inflation artifacts; they're the price of pushing matter toward the level of atomic precision demanded by Moore's Law. Together, these forces create recurring, high-stakes races unmatched by any other market in the world. In these races, firms must commit massive resources years in advance, under extreme uncertainty."

Symposia

Melissa S. Kearney and Luke Pardue have edited a collection of six essays on *Advancing America's Prosperity* (Aspen Economic Strategy Group, 2025, https://www.economicstrategygroup.org/publication/advancing-americas-prosperity/). As one example, Jeremy Neufeld notes in "Aligning High-Skilled Immigration Policy with National Strategy": "Consider: mRNA vaccine technology was developed in the United States because biochemistry pioneer Katalin Karikó was allowed to come to the country in 1985, before the institution of rules in 1990 and 1998 that would have made it much less likely for her to have been able to successfully immigrate. In the years since, 5G was developed in China because Huawei was able to commercialize the research of Erdal Arikan, the Turkish scientist whose breakthroughs on polar codes provided the basis for the technology. Arikan was an international student who studied in the United States, graduating from the California Institute of Technology and the Massachusetts Institute of Technology. He wanted to stay in the United States and only returned to Turkey when he could not secure a green card. Had he faced the immigration system that Karikó faced, he would be a proud American citizen today. Immigration policy thereby seeded a vaccine revolution. Today's immigration policy exported 5G to Shenzhen." As another example, Craig Garthwaite and

Timothy Layton discuss "Coverage Isn't Care: An Abundance Agenda for Medicaid": "Medicaid is now one of the largest public health-insurance programs in the world, enrolling more people than Medicare and more people than the public, social health-insurance programs of the United Kingdom, Germany, or France. . . . The contemporary Medicaid population involves children, pregnant women, the disabled, dual-eligible seniors, and those needing long-term care. Medicaid both pays for 41 percent of births in the US and is the largest single payer for long-term care services in the US. It is the nation's only true cradle-to-grave insurer. . . . Despite this fact, the program largely takes a one-size-fits-all approach . . . "

The *Review of Industrial Organization* presents a symposium on competition policy in Brazil, China, Egypt, India, and five countries of Central and Eastern Europe (October 2025, https://link.springer.com/journal/11151/volumes-and-issues/67-3). Russell Pittman writes the "Editor's Introduction to the Special Issue on Competition Law Enforcement in Developing Countries": "There are at least 125 countries and jurisdictions in the world with competition laws—perhaps more. . . . There were only 12 competition law regimes worldwide in 1970, and two of these—the Japanese and German—were forced upon the losers by the victors in World War II. Countries with market economies gradually adopted competition laws in the post-war period, to the point that there were about 40 competition laws by the time of the fall of the Berlin Wall (1989). Over the 20 years that followed, the number exploded, to at least 110 by 2010; and, according to one authoritative source, there are 135 today: 129 countries and 6 regional organizations. What accounts for this explosion? In the most economically advanced of the Central and Eastern European (CEE) countries, reformers had included competition laws in their legal and regulatory agendas from the beginning. In other CEE countries, the desire for an invitation to membership in the European Community clearly played more of a role . . . More broadly, many developing countries found that loans from the World Bank or the International Monetary Fund, as well as bilateral and multilateral trade agreements with wealthier countries, were conditioned on the writing and implementation of competition laws. . . . [D]eveloping countries around the world have received technical assistance in writing and enforcing competition laws and training agency staffs from sources as diverse as the U.S. Department of Justice, Antitrust Division, and U.S. Federal Trade Commission, the European Commission, OECD, UNCTAD, the World Bank, and successful younger competition agencies such as the Hungarian Competition Authority, the Japanese Fair Trade Commission, and the Korea Fair Trade Commission."

The Bank of International Settlements offers four papers in the *BIS Quarterly Review* based on its 2025 BIS Triennial Survey of Central Banks regarding foreign exchange markets (December 2025, https://www.bis.org/publ/qtrpdf/r_qt2512.htm). For example, Wenqian Huang, Ingomar Krohn, and Vladyslav Sushko contribute "Global FX markets when hedging takes centre stage." "The average daily turnover in over-the-counter (OTC) foreign exchange (FX) spot and derivatives transactions reached $9.5 trillion in April 2025, more than a quarter higher than in April 2022. . . . Over the past three decades, average daily FX trading volumes have not only dwarfed daily global GDP and international trade but also

expanded at a much faster pace. While global FX trading volumes were 12 times global GDP in 1992, they were 30 times that in 2025. . . . They were about 70 times the volume of global trade in 2025, roughly double the ratio in 1992. This rise in FX volumes reflects financial deepening and greater involvement of non-bank financial institutions (NBFIs) in FX markets, as trading for financial motives has come to dominate trading for the purpose of goods and services exchange."

More on Power in Labor Markets

For those who would like more on labor market power than the four papers in this issue of *JEP*, Dan Andrews and Andrea Garnero discuss "Five facts on non-compete and related clauses in OECD countries" (OECD Economics Department Working Papers, April 25, 2025, https://www.oecd.org/en/publications/five-facts-on-non-compete-and-related-clauses-in-oecd-countries_727da13e-en.html). Here are their five facts: "First, the prevalence of such clauses is high, or at least higher than most would have thought, and may be rising. Second, post-employment restrictions have spread to many parts of the economy, well beyond the relatively narrow group of highly paid managers and professionals. Moreover, such clauses appear to be included in employment contracts even where they are legally unenforceable, simply to deter workers. Third, non-compete clauses are most often bundled with other clauses, typically a non-disclosure agreement. This helps to further suppress workers' bargaining power without necessarily protecting trade secrets any more than a non-disclosure agreement alone. Fourth, even when unenforceable, non-compete clauses have economic consequences in terms of reduced job mobility, wages, knowledge spillovers and market dynamism. Finally, apart from the United States, other OECD countries have introduced limitations on the use of such clauses in the last decade, and a few others are currently discussing possible restrictions."

The *Oxford Review of Economic Policy* offers a twelve-paper essay on "New Directions in Competition Policy" (Winter 2024, https://academic.oup.com/oxrep/issue/40/4). Joel Kariel, Jakob Schneebacher, and Mike Walker focus on "Competition policy and labour market power: new evidence and open questions." From the abstract: "We survey existing competition enforcement in labour markets and the empirical evidence on the extent and impact of labour market power, with a particular focus on the UK. We find that in contrast to the US, labour market power in the UK has not risen substantially. Nonetheless, workers vary in their exposure and, for some, labour market power has significant economic costs. Labour market power also appears to interact in significant ways with other labour market policies. We argue these findings underscore the danger of making policy decisions based on evidence from other countries, or by analogy with product markets."

Eric A. Posner and Ruth Zheng recap and follow-up on "The Silicon Valley No-Poach Conspiracy" (University of Chicago, Coase-Sandor Institute for Law & Economics Research Paper Series, 25–18, https://chicagounbound.uchicago.edu/law_and_economics/1033/). From the abstract: "The Silicon Valley no-poach

conspiracy is the most important cartel no one has heard of. It is rarely discussed in the cartel literature and is lost to public memory. More than forty tech firms, including Apple and Google, agreed not to poach employees from one another over three decades, causing an estimated $3.1 billion in lost wages. The Justice Department discovered and broke apart the cartel in 2010, but did not punish the cartel members, who quickly settled with employees and never admitted guilt. However, the case foreshadowed, and perhaps helped spur, two major developments in antitrust law a decade later: the rise of labor antitrust and the erosion of the tech companies' aura of antitrust invincibility."

Discussion Starters

A team of Federal Reserve economists—Matteo Crosignani, Jonathan Kivell, Daniel Mangrum, Donald Morgan, Ambika Nair, Joelle Scally, and Wilbert van der Klaauw—provide evidence on "Financial Inclusion in the United States: Measurement, Determinants, and Recent Developments" (*Economic Policy Review*: Federal Reserve Bank of New York, September 2025, 31:3, https://www.newyorkfed.org/research/epr/2025/EPR_2025_financial-inclusion_crosignani). "The FDIC defines an individual as 'unbanked' if no one in the household has a checking or savings account with a bank or credit union. . . . Among those with a bank account, the FDIC further defines as 'underbanked' those individuals who are banked but underserved by existing saving, credit, and financial products. . . . Underbanked individuals usually pay high fees for accessing their money and for transactions while having few opportunities to build savings and assets. According to the latest 2021 FDIC National Survey of Unbanked and Underbanked Households, there are about 5.9 million households (15.6 million adults) that are unbanked, while 18.7 million households (51.1 million adults) are underbanked. . . . [T]hose who are unbanked accounted for 4.5 percent of U.S. households in 2021, with a further 14.1 percent of households being underbanked."

M. Scott Taylor investigates "Saving Killer Whales Without Sinking Trade: A market solution to noise pollution" (Property and Environment Research Center (PERC), September 25, 2025, https://www.perc.org/2025/09/25/saving-killer-whales-without-sinking-trade/). From the abstract: "Despite generating enormous benefits, maritime shipping has raised the underwater ambient noise levels in the world's oceans by three to four decibels per decade. . . . Drawing on a unique natural experiment, I link changes in the health of a killer whale population to changes in commercial vessel traffic in their critical habitat. Killer whale births are lower, and deaths higher, in noisy years compared to quiet ones—providing the first empirical evidence linking variation in noise from shipping to the health of a marine mammal population. . . . I propose a market-based solution that prices underwater noise pollution that provides a win-win solution for both commerce and conservation."

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary access to JEP articles, go to the AEA website: http://www.aeaweb.org. Change of address notice must be received at least six weeks prior to the publication month.

The Journal of
Economic Perspectives

Winter 2026, Volume 40, Number 1

Symposia

Feature

AMERICAN ECONOMIC ASSOCIATION